Multidimensional analysis in characterizing gastric pathologies by non-invasive indicators

Oksana Lavrenchuk1, Andriy Kharchenko2

1 Institute of Pediatrics, Obstetrics, and Gynecology of the National Academy of Medical Sciences of Ukraine, 8 Platona Maiborody St, Kyiv
2 Department of Chemistry, University of Georgia, 140 Cedar St, Athens, GA, USA

Abstract. The connective tissue is the key medium maintaining the basic homeostatic parameters of the body, taking a direct part in physiological and adaptive processes of the organism. Changes in connective tissue affect the course of gastrointestinal tract pathologies as an organ containing the greatest amount of collagen fibers. In this work, we investigate the pre-epical barrier of the stomach by an array of 13-dimensional indirect non-invasive indicators of primary pathological reactions: immunoglobulins, lysozyme, antimicrobial peptides, glycosaminoglycans, sialic acids, and fucose. We apply discriminant analysis, factor analysis, and principal component analysis to characterize a genetic interference in gastric diseases in order to match each marker with the biological basis of the diagnoses and to identify a diagnosis. We apply a principal component representation to describe each diagnostic group by a non-trivial profile of the principal component features. We show that diagnosing can be implemented via low-dimensional categorization by the quadratic discriminant analysis. We consider multidimensional configurations and apply projection on a factor basis with a maximum likelihood method to improve categorization accuracy. As a result we propose a biochemically conditioned basis for reducing the indicator dimensionality helpful in optimizing the clinical laboratory practice.

Keywords
children, collagen IV, connective tissue disorder, discriminant analysis, factor analysis, gastroduodenitis, principal components

1 Introduction

Chronic gastroduodenitis (CGD) in children prevales in the digestive system diseases especially among adolescent urban population. For example, the frequency of chronic gastritis and HGD in Ukraine is 31.09%, and in Kiev - 42.43% [1]. Late diagnosis, torpidity, susceptibility to relapses, lack of effectiveness of existing prevention activities of CGD is often due to disorders of connective tissue (CTD). The latter fact motivates the importance of exploring the associated pathology - CGD on the background of CTD. However, the clinical practice ready results are rather discursive and insufficient nowadays including diagnostic criteria and protocols of express-assessment. There is also lack of awareness as to the detrimental role of collagen degradation and transformation disorders in the formation of CGD especially in children clinical practice [2].

This work aims to evaluate the information features of biomarkers of the gastrointestinal mucosal barrier in children having the CGD both with and without disorganization in connective tissue. The input for this research is measurements available from noninvasive laboratory methods for indicators of particular components of mucins such as glycosaminoglycans (GAG), sialic acids (SA), and fucose as well as of immunological parameters such as secretory immunoglobulins, lysozyme, HBD-2 in saliva and coprofiltrates (cf) accompanied by biopsy morphologic results. The emphasis on non-invasive evaluation of indirect biomarker indicators of the gastric diseases is made to make results applicable to results produced by general clinical laboratories. The information features are then utilized in categorization of measured indicators based on small training sets of a relatively large dimension, corresponding to a typical laboratory batch.

Defensins (antimicrobial peptides) are the oldest peptides participating in the non-specific protection of the body from infectious agents. To date, three types of defensins are known - α, ß and θ. Only the first two species are found in human body, and the antibacterial defense of the gastrointestinal tract involves ß-defensins which have tissue-specific distribution, expressed by different types of epithelial cells. ß-defensins are binding elements between the innate (non-specific) and the acquired (adaptive, specific) immunity. In the gastrointestinal tract, the β-defensine-2 (HBD-2) is most expressed.
Lysozyme is the most reliable biomarker of nonspecific humoral immunity, a reliable immunological marker for functional state of the gastrointestinal tract.

Immunoglobulins (IgA, slgA and IgG) are markers of a specific immune response, based on the production of plasma IgA - a major factor of local immunity prevailing in all the secretions of the gastrointestinal tract and in the proper plate of its mucosa. The main function of IgA is the detection of antigen in the intestine, preventing the adhesion of bacteria and viruses to the mucous.

slgA is an important indicator of local immunity, formed at the connection of epithelial and immunocompetent cells and resulting from the penetration of Ig from plasma cells of the submucosal space of intestinal epithelium.

Determining features of protective processes on the local level requires determining the concentration of IgG which is present in the largest number in the extravascular channel and is mainly produced via the secondary immune response, playing the role of neutralizer of toxins and viruses, as well as of activation complement system strengthening phagocytosis. IgG binds to cells through the interaction of Fc-receptors. The best-studied Fc-receptors are FcRn and FcγRII that bind IgG, Fcalt epsilon-R, which in turn binds slgE and plgR105, which in turn binds polymeric IgA and IgM. FcRn and plgR perform the Ig transportation while FcRn additionally controls half-life of IgG in serum. FcyR and FceR are signal receptors that, after binding IgG and IgE, regulate various forms of endocytosis, including phagocytosis, antibody production, and inflammatory responses.

GAG and proteoglycans as an obligatory component of the intercellular matrix playing an important role in intercellular interactions, formation and maintenance of cellular forms and organs. Proteoglycans and GAG specifically interact with collagen, elastin, fibronectin, laminin and other proteins of the intercellular matrix. The revealed increase in GAG level in all the patient groups with clinical manifestations of CTD confirms the existence of a dynamic destruction of the biosynthesis of collagen and other proteins of the intercellular matrix in these groups of patients in the direction of prevalence of its degradation processes.

As a result of the catabolism of the components of the main substance, under the action of some hydrolases, the cleavage of neurominidase from the glycoproteins of N-acetylneuraminic (sialic) acid takes place causing the preliminarily destabilized glycoprotein to be absorbed by macrophages. That is why one of the main markers of the extracellular matrix pathology is change in the concentration of fucose and SA in the physiological fluids of the organism (saliva, cf, blood, etc.) describes the connective tissue's state and the progress made by the inflammatory destructive process.

Fucose is not only energy for cells, but also a temporary competitive inhibitor for lectins of pathogens which reduces the load on the mucin layer. It normalizes the microbiota, affects the expression of microbial metabolic pathways, and reduces the expression of bacterial genes of virulence. Rapid fucosing is a protective mechanism that uses macroorganism resources to maintain interactions with the microbes during pathogen-induced stress.

Thus, an increase in the level of fucose in cf in children reflects the degree of connective tissue destruction and indicates the withdrawal of glycoproteins in inflammatory processes in the upper gastrointestinal tract on the CTD background. Reducing the amount of SA indicates a destabilization of protective processes of mucous membrane and their reduced endurance under the influence of damaging factors. The latter increases the likelihood of the development of destructive processes in patients with CTD indications.

The CGD triggers changes in the individual immunological and biochemical components of the mucociliary layer based on the unifying effect of inflammation. An increase in HBD-2, lysozyme, immunoglobulins (slgA, IgA, and IgG), fucose, and SA is expected as a result of an inflammatory processes in the upper gastrointestinal tract [3]. With the development of gastroduodenitis over CTD, local components of the mucous layer will be gaining features of deviation so a reverse picture will be expected: the level of the molecular marker HBD-2, Ig, lysozyme and SA lower than those not only for HHD, but also for healthy patients. The DST causes a significant change in the composition of individual biochemical constants in mucous membrane and proteoglycans that are part of cell membranes and mucous secretion. Increasing the level of GAG in individuals with clinical manifestations of GST indicates the existence of a dynamic destruction of the biosynthesis of collagen and other proteins of the intercellular matrix in the direction of further degradation. High levels of fucose and low SA in CTD patients reflect the degree of connective tissue destruction and indicate the intensity of glycoprotein withdrawal in inflammatory processes.

Our goal is four-fold. First, we need to determine the utility of the non-invasive biomarkers in covering measurement result distribution of the 4 related patient groups. Particularly we are interested in examining their redundancy within each diagnose. Second, we need a reliable characterization of all these diagnostic groups stable over fluctuations of lab equipment and choice of commercially available chemical sets used available at multiple laboratories of mixed research and clinical levels. Third, we need to examine the most difficult diagnosis – CGD with CTD – versus main factors, genetic and pathogenic, and check if the corresponding data is representable in the 2-factor basis or we are dealing with a more complex probably non-linear phenomenon. Four, our other aim is finding informative profiles for each diagnostic group along the lines of similar efforts of other authors [4, 5].
Materials, methods, and data

Given 4 groups of patients with diagnoses known from biopsy and other invasive methods – a group having a genetically predetermined CTD (group 3), a group undergoing pure H.Pylori-based inflammation (group 2), a group undergoing H.Pylori-based inflammation on top of genetically predetermined CTD (group 1), and a group of healthy individuals (group 4) – and data of biomarker responses obtained from non-invasive methods.

The state of local immunity was determined by a study of specific and non-specific immunity factors in local secrets (saliva, cf). Humoral factors of local immunity (sIgA, IgA, IgG) in cf and saliva was measured by radial immunodiffusion in gel. HBD-2 concentration was measured by immunoassay analysis with β-Defensin 2 ELITA Kit by Immunodiagnostik AG. Concentration of lysozyme in saliva and cf was determined using the dry powder of one-day culture by Micrococcus Lzychodeicus. Concentrations of GAG, fucose (methylpentose), and SC per gram of protein in salivary and cf were determined in by the Lowry method. Endoscopic gastric biopsy in children was received in accordance with the Universal Declaration on Bioethics and Human Rights, accepted by UNESCO General Conference on October 19, 2005.

Endoscopic biopsies were fixed in a 10% solution of neutral zubufernogo formallin for the determination of morphological characteristics. The mucus-forming function of the epithelial cells was investigated histochemically via PAS-reactions with schiff-iodine acid. Immunohistochemical features of biopsies with a study of Collagen Typ IV content were determined by the indirect streptavidin-peroxidase method. Interpretation of the immunosuppression results was performed using monoclonal antibodies to Collagen type IV.

Pathomorphological, histochemical and immunologist-chemic studies of biopsy specimens were performed on Axioskop 40 microscopes and Olympus CX with a computer digitizer. Measurements results were processed using the sklearn.decomposition, R, and MATLAB software packages. Statistical analysis included analysis of frequencies and contingency tables, Student's t-criterion, ϕ* criterion, the χ2 test with the Yates correction, G-test, Welch T-test, Mann-Whitney U-criterion, and Fisher exact test. Multidimensional datasets were examined with factor analysis method, linear and quadratic discriminant analysis. Principal component analysis with the rotation with varimax and promax basis rotation [6] was applied for feature extraction. The lab measurement batches processing and training set control initiative has being actively developed as a Gunicorn/Flask/nginx service under OS Ubuntu.

Discussion

Judging only by the levels of biomarker expression, the categorization problem can be trivial or hard. For example, the expression of sialic acids in cf in all the groups is intuitive: the pure inflammation group’s level (#2) is higher than in the healthy one (#4), presence of a congenital inter-cellular matrix disorder causes an inhibited expression (group #3), and a picture of the mixed case of an inflammation of a degenerate tissue (group #1) would produce an interpolation of the above cases like in Figure 1, left.

Figure 1. Example of biomarker expression in all the diagnostic groups. Left – quite expected layout of distributions of immunoglobulin G levels, right – way less intuitive layout of sialic acids levels.

The reality with most other markers is the opposite like in Figure 1, right. Thus, shattering marker expression vectors between groups #1-4 univariately i.e. using a system of thresholds is not feasible due to the above obvious non-linearity. A
bivariate categorization via a LDA or QDA producing a set of equations of boundaries between classes looks quite promising – see Figure 2.

![Figure 2](image)

**Figure 2.** β-defensin and secretory immunoglobulin A in cf as a DA basis providing the total categorization accuracy 78.57% via QDA model

\[-8.457 \times -0.129y - 2.77x^2 + 0.378xy - 0.009y^2 = -6.578, 1.254x + 0.183y + 3.469x^2 + 0.378xy - 0.009y^2 = -0.288, 7.066x + 0.101y + 159.005x^2 - 0.495xy - 0.008y^2 = 1.848, 9.712x + 0.0535y + 6.239x^2 - 0.214xy - 0.008y^2 = 6.29, 15.523x - 0.028y + 161.775x^2 - 0.874xy + 0.001y^2 = -1.848, 5.811x - 0.081y + 155.536x^2 - 0.6xy + 0.0008y^2 = -1.56\]

where \(x\) and \(y\) are measured levels of sIgA and HBD2 in cf respectively. Crossed out points denote categorization errors.

While producing a relatively low error rate, a critical limitation of this approach is fluctuation in the base level between lab test batches and between participants of a research collaboration. A workaround may be utilizing a non-trivial layout of biomarker levels’ correlation.

Our optimism about the possibility to resolve all the 4 diagnostic groups via distinct correlating marker layouts can be tested by a FA of the future training dataset versus increasing number of factors: the dataset projected on the correct number of factors should result in better categorization compared to the best accuracy obtained from the bivariate DA experiments. Our numerical experiments confirm the existence of 2 major underlying factors in the data – expectedly, one of the genetic predisposition and the other related to inflammation. See Figure 3.

![Figure 3](image)

**Figure 3.** Maximum likelihood 2-factor analysis of a typical lab batch of 64 patients (right) and QDA in the basis of the latent factors providing 86.51% total accuracy (left).
Although FA, as we know aiming at maximizing between-class distance, in principle confirms our hypothesis about the 2 factors governing the manifold of measurement result values, we need a way to access the within-class correlation to estimate the non-invasive biomarkers’ redundancy and response to the factors. In simple words, agreement between vectors is easy depictable in the form of linear correlation therefore we will apply the principal component analysis to examining the data. As with the FA, the whole measurement vectors’ projection on principal components is expected to provide an improved QDA categorization.

Clearly enough, receding compactness of diagnosis clusters in the basis of principal components 2 and 3 confirms our initial hypothesis about the 2 latent factors. Since PCA is an instrument of analysis of distributions, let us examine correlating subsets of indicators: group A – highly correlated fucose and GAG in cf; group B - highly correlated IgA and SA in cf as well as HBD2 in saliva; group C – the rest indicators. This grouping can be used to reduce the dataset dimensionality and extract features for further categorization. As each diagnostic group corresponds its own biologic background, we can further examine the principal component layout within each group – see Figure 4.

**Figure 4.** PCA results of individual diagnose groups illustrating non-trivial correlation profile of each group.

We observe that members of the control group exhibit independence of all the indicators in the basis of principal components 1 and 2 except salivary sIgA, IgA, GAG, and CK in cf. Patients of the CGD over CTD (group 3) feature a relatively high correlation between indicators: fucose, SA, and IgA in cf. These indicators are in anticorrelation with sIgA in cf as well as IgG, sIgA, and GAG in cf that confirms a hybrid nature of this group.

The CGD (group 2) group features clinically informative 2 subsets of strongly correlating indicators: {IgA, IgG in cf}; {GAG and sIgA in cf}. Lysozyme has less correlation significance in cf, and independently of them there is an independent indicator - HBD-2 in cf. The correlation between the pairs of HBD-2 and sIgA sIgA indices in saliva and IgA and IgG in saliva is not very pronounced. A strong mutual anticorrelated dependence (salivary HBD-2, IgA, sIgA, and IgG) are observed between them. We note that in this group the only principal component is parallel with HBD-2 in cf.

The group of children with DST (group 1), the following picture of the correlating subpopulation of parameters was observed: subgroup A (lysozyme, IgA and SC in cf); Subgroup B (fucose in cf and salivary sIgA). The lack of anti-correlation pairs is a property of this group. Fucose is the weakest of all correlates with PC 1 and 2, obviously because the metabolic function of CT in this diagnostic group is disturbed. The correlation of PC 1 or 2 with GAG in cf belongs to all 3 diagnostic groups of patients, which illustrates metabolic processes in ST. It is also no coincidence that correlations with PC 1 or 2 fucose, which determines the degree of destruction of the mucous layer.

Indicators in healthy children (group 4) are featured by high mutual independence and anti-correlation of salivary IgG with GAG, SC and IgA in cf.
Thus, based on measurements of immunobiochemical parameters in patients with a simple (CGD) and combined (CGD over CTD) pathology, it can be stated that there are at least two factors that ensure the variability of measurement values in each patient group: the acquired inflammatory factor and the factor of the subsequent violation of the integrity of CT. For the purpose of studying the second factor, an identical immunobiochemical study of healthy children from the point of view of CGD was performed (hypermobility of joints, skin hypersensitivity, anomaly of the spine, eye abnormalities, high palpation, flatulence, etc.).

The results obtained indicate that immunobiochemical disorders of the mucous-cellular layer in children with CGD over CTD and without CTD differ significantly in certain indicators. Analyzing the indirect effect of CTD on a number of non-invasive indicators used in this paper, we departed from the known fact that the key CT component, Collagen type IV, is associated with the clinical picture of CTD. Deviations in the formation mechanism of Collagen BM and microfilaments are critically important for all the processes in stomach. Anomalies of BM homeostasis directly lead to certain immunogenic response. It should be noted that at the physiological level, the root cause of this immunogenic response is the process of formation of structurally different forms of collagen due to the violation of its metabolism.

Conclusion

Using a 2-factor segmentation of a manifold of gastric diseased patients including a subset of healthy ones and a (potentially redundant) set of biomarkers covering the biological background of possible pathologies let us find correlation-based indicator profiles helpful in optimizing the current clinical laboratory practice.

We draw several features of each diagnostic group. Within the CGD over CTD, the highest accuracy is provided by: HBD-2 in cf and salivary IgG, HBD-2 and IgA in cf, HBD-2 in cf and CK in cf. The following configurations provided the total accuracy as high as 93.9%: salivary HBD-2 and lysozyme in cf, salivary IgA and HBD-2 in cf, salivary IgG and SA, salivary IgA with lysozyme in cf, salivary HBD-2 with lysozyme in cf, salivary HBD-2 and lysozyme in cf, salivary IgG and SA. The strongest anticorrelation and SA in cf, that signals about malfunctioning extracellular matrix of the stomach as well as about a long-term disrupted immune defense and changes in the mucin layer is emphasized with correlating indicators from another subset the most prominent that is an indication of disrupted humoral link in the low segment of the stomach. The whole picture of the CGD group, one of the markers of mucous membrane destruction is a complex of correlating indicator clusters: fucose correlating with SA and IgA in cf, and IgG and GAG in cf. The strongest anticorrelation connection was observed between slgA and IgG in cf. An existence of these anticorrelating subgroups is an indication of disordered mucin layer that in its turn depends on the influence from the anti-infection barrier.

In the CGD group, patients’ defensive mechanisms crank up only after pathogens entering the mouth. This is pointed out by an anticorrelating pairs salivary slgA and IgA, and IgG and HBD-2. In this group, the subgroup IgG and IgA in cf is the most prominent that is an indication of disrupted humoral link in the low segment of the stomach. The whole picture of disrupted immune defense and changes in the mucin layer is emphasized with correlating indicators from another subset GAG and slgA in cf, that signals about malfunctioning extracellular matrix of the stomach as well as about a long-term action of an irritating agent. The latter can be confirmed by discovering lysozyme in cf and an increased level HBD-2 in cf.

The CTD group is featured by the absence of any prominently anticorrelating pair of indicators. This in general points to unavailability of the stomach cell protective mechanism. These patients can be attributed to the risk group with uncontrolled episodes of immune system failures.
References


