



Short Communication

Total IgG method insensitivity to the presence of IgG4 immunoglobulins

Q1 Andre Mattman ^{a,*}, Sophia Wong ^b, Nadine Urquhart ^a, Patrick Wong ^c^a Department of Pathology and Laboratory Medicine, St. Paul's Hospital, University of British Columbia, Vancouver, Canada^b UBC Medical Biochemistry Residency Training Program, University of British Columbia, Vancouver, Canada^c UBC General Pathology Residency Training Program, University of British Columbia, Vancouver, Canada

ARTICLE INFO

Article history:

Received 28 January 2014

Received in revised form 27 March 2014

Accepted 28 March 2014

Available online xxxx

Keywords:

IgG4 related disease

IgG4

Antigen excess

Hook effect

Analytic error

Quality control

ABSTRACT

Objective: IgG4 related disease (IgG4RD) is detected in part by measurement of serum IgG4 levels in patients suspected to have the condition. The immunonephelometric methods that are used to measure serum IgG4 are susceptible to hook effect errors. The quality control parameter IgGRDiff = $100 \times [\text{Sum (IgG1, IgG2, IgG3, IgG4)} - \text{Total IgG}] / \text{Average [Sum (IgG1, IgG2, IgG3, IgG4), Total IgG]}$ is theoretically of use in detecting large IgG4 hook effect. We evaluated the IgG4RDiff parameter in this context.

Design and methods: Retrospective review of the IgGRDiff parameter in a cohort of patient samples. All samples were tested with the Siemens BNII analyzer utilizing Siemens (total IgG) and Binding Site (IgG subclass) reagents. The 95th percentile for the absolute value of IgGRDiff was determined and compared to the IgGRDiff values for the samples with known hook effect error. IgGRDiff was regressed against IgGRsum [IgG4Rsum = $100 \times \text{IgG4} / (\text{IgG1} + \text{IgG2} + \text{IgG3} + \text{IgG4})$].

Results and discussion: The 95th percentile of the absolute value of the IgGRDiff parameter was 22.8% in our study sample set ($n = 163$). The absolute value of the IgGRDiff was <22% in 4/6 cases indicating low sensitivity of IgGRDiff in identifying IgG4 hook effect errors. IgG4Rsum was directly correlated with IgGRDiff (IgGRDiff = $1.21 \times \text{IgG4Rsum} - 7.9$, $R^2 = 0.62$) indicating that the total IgG measurement has little or no reactivity with IgG4 immunoglobulins.

Conclusions: The Siemens' Total IgG method is insensitive to IgG4 and should not be used as a quality control parameter to detect IgG4 hook effect.

© 2014 The Canadian Society of Clinical Chemists. Published by Elsevier Inc. All rights reserved.

Introduction

The evolving understanding of IgG4 related disease (IgG4RD) has led to a number of recent publications on all aspects of this treatable syndrome, including the diagnostic process [1–7]. The diagnostic process remains hinged on histopathology [6] but blood tests have an important role in clarifying the likelihood of disease, particularly in the case of IgG4RD manifesting as autoimmune pancreatitis [3,4,7]. In this regard, two recent publications [8,9] have commented on laboratory quality control aspects associated with the most useful blood test, serum IgG4, which has a relatively high sensitivity albeit a low

specificity for the diagnosis of this disorder [2,5]. Specifically, Khosroshahi et al. [8] commented on a major limitation of the IgG4 test – the hook effect – which can variably affect some laboratory immunonephelometric methods [8,10], requiring vigilance by laboratory specialists and clinicians alike. Relatedly, McLean-Tooke et al. [9] reported on the utility of measuring IgG subclasses and total IgG independently as a means to identify samples affected by the hook effect with regard to any of the individual IgG subclass measurements, including IgG4. To this end, McLean-Tooke et al. statistically characterize a calculated parameter (see Eq. (1)) in a series of patient samples noting a variance in IgGDiff that is predicted not by the presence of IgG4 hook effect errors, but by IgG subclass composition.

$$\text{IgGDiff}(\%) = 100 \times [\text{IgGsum} - \text{IgGtot}] / \text{IgGtot}. \quad (1)$$

IgGsum IgG1 + IgG2 + IgG3 + IgG4, where the subclasses are measured individually in g/L

IgGtot total IgG, where IgG is measured in g/L with the use of polyclonal antibodies intended to recognize all IgG subclasses

With regard to the detection of IgG4 hook effect errors, we have noted that the IgGDiff parameter has apparent limitations as a quality

Abbreviations: IgG4RDis, IgG4 related disease; IgGtot, IgG total measured using polyclonal antibodies that recognize all IgG subclasses; IgGsum, Sum of the individual IgG subclass measurements; IgGDiff, percentage difference between IgGtot and IgGsum normalized by IgGtot; IgGRDiff, equivalent to IgGDiff but uses average of IgGtot and IgGsum as the denominator; IgG1Rsum, percentage of IgGsum composed of IgG1; IgG2Rsum, percentage of IgGsum composed of IgG2; IgG3Rsum, percentage of IgGsum composed of IgG3; IgG4Rsum, percentage of IgG4 composed of IgGsum.

* Corresponding author at: Department of Pathology and Laboratory Medicine, St. Paul's Hospital, 1081 Burrard St., Vancouver, BC, V6Z 1Y6, Canada. Fax: +1 604 806 8815.

E-mail address: amattman@providencehealth.bc.ca (A. Mattman).

control measure in our laboratory. Our laboratory has utilized a specific set of commercial test reagents and instrumentation over the past four years that is distinct from that employed in the McLean-Tooke et al. study. We follow a quality control monitoring system for IgG subclass measurements as outlined in the materials and methods section. In this time, we have observed the release of multiple IgG4 hook effect errors even while monitoring the IgGDiff parameter [10].

The release of multiple IgG4 hook effect errors while monitoring the IgGDiff parameter is unexpected. In the setting of IgG4RD, patient samples often have an absolute elevation in serum IgG4 concentrations, as well as a relative increase in the IgG4 level as a percentage of IgGtot or IgGsum (IgG4Rsum) [2,5]. Given the increase in IgG4Rsum associated with IgG4RD, one expects the IgG4RD parameter to be a useful indicator of the hook effect error, to which IgG4RD patient sample measurements are susceptible. Specifically, a large hook effect error of IgG4 would theoretically lead to a marked negative bias in IgG4 measurement, which would be paralleled in a negative bias of IgGsum, and IgGDiff. Thus, the IgGDiff parameter should be a sensitive indicator of a large hook effect error affecting IgG4 measurement.

However, this proposed utilization of the IgGDiff parameter may not be as effective as expected in practice owing to calibration error that affects some test methods. In particular, McLean-Tooke et al. [9] noted that the relative difference between total IgG measurements and the sum of individual IgG subclass measurements was modulated by relative IgG subclass proportions i.e. calibration error of the total IgG measurement may limit its utility as a quality control parameter. Any potential calibration error affecting an IgGtot method may be specific to a particular brand of commercial reagents. We are unaware of similar evaluations having been performed on the Siemens reagents in use in our laboratory.

In view of this possibility of a calibration error affecting the IgGtot method, which would affect the utility of monitoring the IgGDiff parameter, we determined to study these parameters in our own laboratory. However, in view of possible errors with the denominator of the IgGDiff parameter, IgGtot, as highlighted by the McLean-Tooke article [9], we are now monitoring the related parameter IgGRDiff (see Eq. (2)):

$$\text{IgGRDiff (\%)} = 100 * [\text{IgGsum} - \text{IgGtot}] / \text{Average}[\text{IgGsum}, \text{IgGtot}] \quad (2)$$

Therefore, our investigation was specifically designed to evaluate the characteristics of the IgGRDiff parameter in a series of patient samples previously tested in our laboratory utilizing one brand of commercial reagents for IgG subclasses (Binding Site), another for IgGtot measurements (Siemens), with all measurements performed on one instrument (Siemens BNII nephelometer). We sought to determine the central 95th percentile confidence limits of the IgGRDiff parameter and to determine if its magnitude, in patient samples, was affected by the patients' IgG subclass distributions. The study sample set included 6 patient samples which had previously been reported out with falsely low IgG4 levels due to hook effect error, and 1 sample which had previously been noted to be susceptible to falsely low IgG3 measurement due to a hook effect error. For these samples the IgGRDiff parameter was evaluated both on the original sample results (to evaluate the detection rate for IgG4 hook effect), and on the hook effect corrected results. The sample set also contained patient test results which had been screened for hook effect errors and found to be at low risk.

Materials and methods

Patient results

We conducted a retrospective review of the IgGRDiff parameter in a cohort of patient sample results from samples submitted for the investigation of IgG4 related disease and/or for IgG subclass deficiency. We included only the first set of sample results for each unique patient for a 4-year time period dating back to January 1, 2010. IgG subclass results

that were known to be affected by the hook effect (see Evaluation of patient samples for hook effect errors) were corrected using results from higher dilutions prior to inclusion in the data set, with the exception of the evaluation of the sensitivity of the IgGRDiff parameter for hook effect error detection. In this latter analysis the IgGRDiff was evaluated using the original IgG3 and IgG4 measurements as performed at a 1:100 dilution. Patient samples with known monoclonal proteins, or with abnormally low levels of one of the four subclasses, were excluded from the analysis given that the IgGRDiff parameter may fluctuate greatly under these conditions. All other patient samples were included.

Chemistry reagents

All samples had been previously tested with the Siemens© (Munich, Germany) BNII nephelometer utilizing Siemens total IgG and the Binding Site© (Birmingham, UK) IgG subclass kit LK001 TB with latex-enhanced IgG3 and IgG4 antibodies.

Evaluation of patient samples for hook effect errors

All samples had been assessed for the likelihood of IgG4 hook effect by determination of the IgGtot value, and by evaluation of the IgGDiff parameter. Based on a previous analysis [10], IgGtot values were identified as a strong predictor of the likelihood of an IgG4 hook effect error. If the total IgG indicated a higher risk for the hook effect, a higher initial sample dilution was applied to avoid the hook effect. Specifically, if the total IgG value was < 10 g/L, a standard 1:100 dilution was utilized for the IgG4 determination. If the total IgG was between 10 and 20 g/L, a 1:400 dilution was utilized for the IgG4 determination. If the total IgG was ≥ 20 g/L, then a 1:2000 dilution was used for the initial IgG4 determination. Independently, the IgGDiff parameter was determined in each case and, if ≥ ± 20%, samples were retested utilizing higher dilutions to rule out hook effect errors. The IgGDiff ≥ ± 20% criterion was in place from Jan 1, 2010 forward. In 2011, a retrospective assessment of potential hook effect errors affecting the previously reported (2010) IgG4 tests identified 6 cases of hook effect error. This retrospective review was conducted using leftover samples stored up to 6 months at – 20 °C, and patient recall, as necessary, to retest select samples and or patients from 2010. Sample retrieval, or patient recall, was considered important if the sample met two or more of the following criteria: i) clinical indication for testing was clearly indicated as suspected IgG4RD, or a specific manifestation (eg autoimmune pancreatitis) of IgG4RD, ii) previously reported IgG4 concentration on the standard 1:100 dilution was normal or abnormally low, iii) previously reported total IgG > 11 g/L along with an IgG4:IgGtot < 0.025. This retesting process identified a total of six IgG4 hook effect errors out of 53 total sample retests. From 2011 forward, total IgG based determination of the initial IgG4 dilution was applied prospectively.

Data analysis

The mean and 95th percentile for the absolute value of IgGRDiff parameter was calculated on all samples that met the inclusion/exclusion criteria as defined. The confidence interval from this analysis was used to define a zone of results in which IgG subclass results, as a group, were considered analytically accurate without need of further result verification. The sensitivity of the IgGRDiff parameter for detection of hook effect errors was assessed by measuring its performance in a small series of 7 patient samples that had previously been identified as susceptible to the hook effect. Finally, we sought to further characterize IgGRDiff by comparison of this parameter to the absolute and relative proportions of the individual IgG parameters in the entire pool of results that met study inclusion criteria. Specifically, we performed linear regression analysis of the IgGRDiff parameter against each of the individual IgG subclasses (IgG1, IgG2, IgG3 and IgG4) relative to IgGsum (IgG1Rsum, IgG2Rsum, IgG3Rsum and IgG4Rsum). Finally, stepwise

discriminant analysis with $p < 0.05$ was used to determine which of several variables were predictors of an IgGRDiff value greater than 22%, when using hook effect-corrected IgG subclass values. Variables chosen for inclusion in the discriminant analysis included the absolute value of the individual IgG subclasses (IgG1, IgG2, IgG3 and IgG4) and their respective ratios to IgGsum (IgG1Rsum, IgG2Rsum, IgG3Rsum and IgG4Rsum), IgGsum, IgGtot, the sum of {IgG2, IgG3, IgG4}, and the ratio of the sum of {IgG2, IgG3, IgG4} to IgGsum. All statistical analyses were performed with SPSS, version 19.

Results and discussion

There were 349 IgG subclass and matching IgGtot tests performed on patient samples over the period of study of this retrospective review. Of these, there were 321 unique patient samples. Four of these samples were known to have a monoclonal protein and were excluded from the study. We also excluded 135 patient samples with a IgGtot less than 7.00 g/L and a further 19 patients with an IgGtot greater than or equal to 7.00 g/L but an abnormally low measured level (after correction of known hook effect errors) of one of the individual IgG subclasses. Thus, the final number of samples included in our study was 163 samples.

The IgGRDiff and the IgGDiff parameters were both positively skewed. The median IgGRDiff value was -2.10% with a range of -22.8% to $+52.1\%$. The 95th percentile of the absolute value of the IgG4RDiff parameter was 22.8% . This absolute value of 22.8% was truncated to 22% for use as a screen for potentially erroneous IgG subclass, or IgGtot, measurements in need of verification. The corresponding values for the IgGDiff parameter were a median of -2.10% , a range of -20.3 to $+70.4\%$, and the 95th percentile of its absolute value was 25.8% . The IgGDiff parameter was not further considered in our analysis.

Considering the IgGRDiff threshold of $\pm 22\%$, we next determined how effective it would have been, had it been in use when the 7 samples with known hook effect error had been originally tested using the standard 1:100 sample dilutions. We found that the observed IgGRDiff values for the 6 IgG4 hook effect samples ranged from -41% to $+21\%$ ($-41, -31, 5, 8, 14$, and 21%) when the IgG4 concentration was measured using the standard settings of a 1:100 sample dilution. After sample dilution to as high as 1:8000, the corresponding IgG4 values increased from less than 1 g/L, in each case, to values of 22, 23, 4, 14, 10 and 10 g/L, respectively. This degree of underreporting, theoretically, should have resulted in IgGRDiff values whose value was $< -20\%$ in all cases rather than in only two. In all 6 cases, after appropriate sample dilution, the corrected IgGRDiff parameters increased to $\geq +22\%$ (22, 27, 30, 48, 51, and 52% , respectively) suggesting that the IgGtot measurement might be negatively biased in these specific samples. Thus, while the IgGRDiff parameter is a potentially useful indicator of an IgG4 hook effect, its efficacy appears markedly diminished, possibly due to a large negative bias affecting the IgGtot measurement itself.

With regard to hook effect errors affecting other subclasses, we identified only one sample, in this specific case a hook effect error affecting the IgG3 measurement. In contrast to the IgG4 hook effect cases, this one known case of IgG3 hook effect error was associated with a larger absolute value of the IgGRDiff parameter (IgGRDiff equal to -51%), corresponding to an original IgG3 measurement of 0.06 g/L when measured at a standard 1:100 dilution (no instrument analytic error flags noted). After a higher dilution (1:2000), the true IgG3 value of 8.5 g/L resulted in an IgGRDiff value of 1% . This case of IgG3 hook effect was associated with a polyclonal increase in IgG levels.

After reviewing these results, we next investigated the hypothesis that the IgGRDiff value is relatively insensitive to the IgG4 hook effect, due to calibration error affecting the IgGtot method. Similar to the McLean-Tooke et al. [9], we have noted that the IgGtot method in our laboratory differs systematically from the IgGsum when total IgG increases, with this effect modulated by the relative concentration of the major IgG subclass, IgG1. In this analysis, we specifically looked for an

association between IgG calibration error in relation to IgG4 and other individual IgG subclass values by means of linear regression and stepwise discriminant analysis using the data set of unique patients whose samples met the described inclusion and exclusion criteria for the specificity analysis ($n = 163$ samples).

Examination of the relationship between IgGRDiff and the relative proportion of the respective subclasses indicated that IgGRDiff was significantly affected by IgG subclass composition. Specifically, there was a significant association of IgGRDiff with both IgG1Rsum (see Fig. 1) and IgG4Rsum (see Fig. 2), with the IgG4Rsum correlation being the stronger. In fact, the slope of the curve in Fig. 2 is 1.21 (95% CI 1.06 to 1.35) with an R^2 value of 0.615 (95% CI 0.512 to 0.699 , $p < 0.001$) indicating that the IgGRDiff parameter is directly related to the IgG4Rsum. In contrast, the regression statistics for IgGRDiff with IgG2Rsum (slope = 0.18 , 95% CI -0.01 to 0.38 , $p = 0.068$; $R^2 = 0.021$, 95% CI 0 to 0.085 , $p = 0.065$) and for IgGRDiff with IgG3Rsum (slope = 0.09 , 95% CI -0.54 to 0.71 , $p = 0.788$; $R^2 < 0.001$, 95% CI 0 to 0.026 , $p = 0.8$) revealed no significant associations. There was also a weak but significant positive correlation between IgGRDiff and IgGtot (slope = 0.34 , 95% CI 0.16 to 0.52 , $p < 0.001$; $R^2 = 0.077$, 95% CI 0.017 to 0.170 , $p < 0.001$).

The strong relationship between IgGRDiff and IgG4Rsum is further illustrated by the fact that 10/11 patients with IgGRDiff $\geq +22\%$ had an IgG4Rsum greater than 16% (normal $< 5\%$). The remaining patient had a relatively low total IgG level (7.1 g/L) and unremarkable individual IgG subclass levels. In stepwise discriminant analysis, the IgG4Rsum was the sole significant predictive parameter of whether IgGRDiff would exceed $+22\%$. When high IgG4Rsum ($> 16\%$) values were removed, only 3/147 (2%) samples had an absolute IgGRDiff $> 22\%$. Altogether, these findings suggest that there is a specific relationship between IgG4 levels and errors in calibration of the total IgG method. Specifically, the IgG4 subclass immunoglobulins appear to have limited reactivity with the reagent antibodies utilized in the Siemens total IgG method.

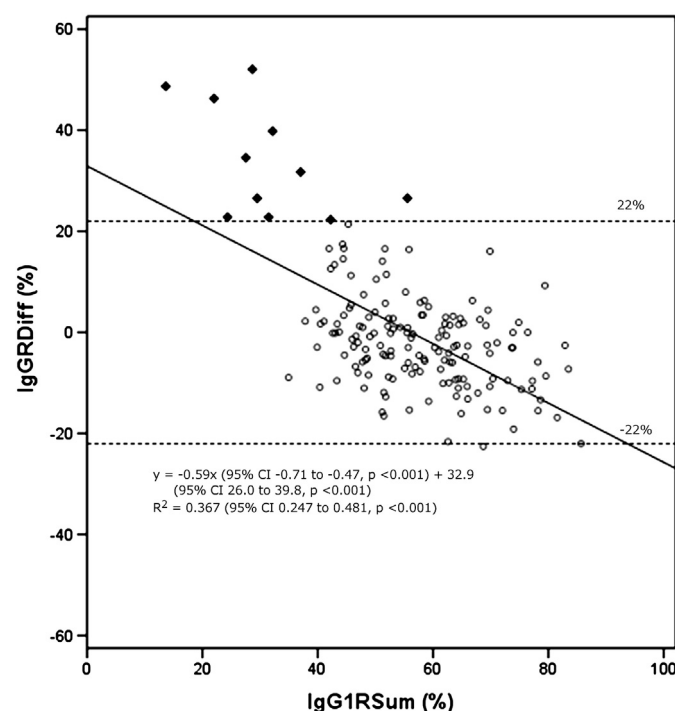


Fig. 1. IgGRDiff vs IgG1Rsum where IgGRDiff is given in Eq. (2) and IgG1Rsum represents the parameter $\text{IgG1}/\text{Sum}(\text{IgG1, IgG2, IgG3, IgG4})$ (%). The dashed lines represent the IgGRDiff $\pm 22\%$ thresholds. The solid diamonds indicate patient test results that have an IgGRDiff $\geq 22\%$. The linear regression equation and R^2 values are provided on the figure.

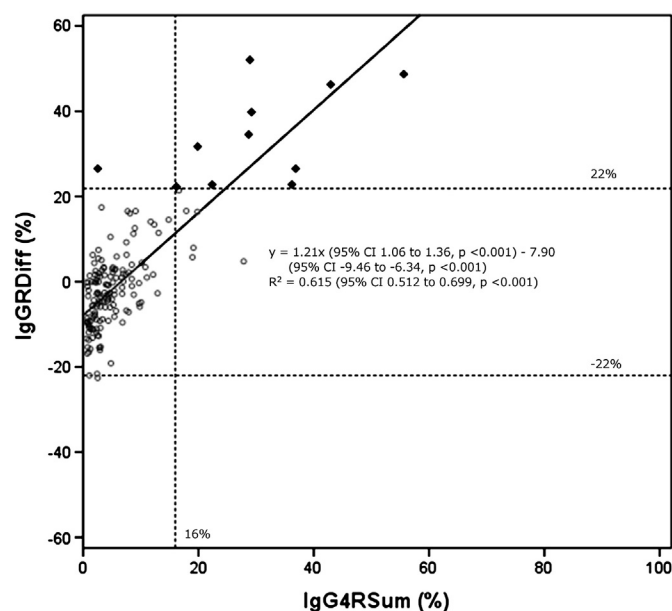


Fig. 2. IgGRDiff vs IgG4RSum where IgGRDiff is given in Eq. (2) and IgG4RSum represents the parameter IgG4/Sum(IgG1, IgG2, IgG3, IgG4) (%). The dashed lines represent the IgGRDiff $\pm$ 22% thresholds. The solid diamonds indicate patient test results that have an IgGRDiff $\geq$ 22%. The linear regression equation and R^2 values are provided on the figure.

Since samples from IgG4 related disease may lead to two concurrent analytical errors, an increased risk of the hook effect (with a consequent negative bias in IgG4), and an underestimation of the IgGtot due to insensitivity of IgGtot to the presence of the IgG4 immunoglobulins, the IgGRDiff (see Eq. (2)) may be relatively close to zero. This interpretation would explain the apparent insensitivity of the IgGRDiff parameter in the detection of IgG4 hook effect errors.

Our evaluation of the IgGRDiff differs importantly from the evaluation performed by McLean-Tooke et al. [9] with regard to reagents used and with regard to statistical analyses but nevertheless show parallels in the sense that total IgG error in both cases was related to relative IgG subclass composition. Other total IgG methodologies may be similarly affected by IgG subclass composition but this further analysis was beyond the scope of this retrospective review. We anticipate a prospective analysis of the effect of IgG subclass composition on total IgG levels using the different manufacturer's total IgG method and IgG subclass methods.

Our study has important limitations. First, we cannot be certain that we have identified all patient samples that were affected by a hook effect error. The detection rate of our screening approach for hook effect errors has not been formally validated in a prospective manner. Second

the evaluation of the sensitivity of the absolute value of the IgGRDiff parameter of 22% as an indicator of IgG4 hook effect errors is on too small a sample set to provide definitive information. Finally, the derived confidence limit for the absolute value of IgGRDiff (22%) applies only to those samples with normal or high levels of each of the individual IgG subclasses, and of IgGtot as well, as these were the criteria for entry of sample data into our analysis.

Conclusions

An IgG4RDiff absolute value of 22.8% represents the 95% confidence limits for this parameter as determined by retrospective analysis of patient sample data analyzed for IgGtot and IgG subclasses. This confidence limit proved to be an insensitive marker for the IgG4 hook effect although the sample size for this sensitivity analysis was too small to provide definitive results. We found that there was a specific positive correlation between IgGRDiff and IgG4RSum indicating poor sensitivity of the IgGtot measurement to IgG4 concentration. This negative bias of the IgGtot test in the setting of elevated IgG4RSum explains the observed poor sensitivity of IgGRDiff to the presence of the IgG4 hook effect. In the absence of hook effect errors, the IgGRDiff parameter is positively skewed by patient samples with relatively high IgG4 concentrations, as these account for 10/11 patient samples with an IgGRDiff $\geq$ +22%. This latter observation is also attributed to the apparent IgG4 insensitivity of the Siemens IgGtot method.

References

- [1] Mahajan VS, Mattoo H, Deshpande V, Pillai SS, Stone JH. IgG4-related disease. *Annu Rev Pathol: Mech Dis* 2014;9:315–47.
- [2] Boonstra K, Culver EL, Wenniger LM, Heerde MJ, Erpecum KJ, Poen AC, et al. Serum IgG4 and IgG1 for distinguishing IgG4 associated cholangitis from primary sclerosing cholangitis. *Hepatology* 2013.
- [3] Sah RP, Chari ST. Serologic issues in IgG4-related systemic disease and autoimmune pancreatitis. *Curr Opin Rheumatol* 2011;23:108–13.
- [4] Patel H, Khalili K, Kyoung KT, Yazdi L, Lee E, May G, et al. IgG4 related disease — a retrospective descriptive study highlighting Canadian experiences in diagnosis and management. *BMC Gastroenterol* 2013:168.
- [5] Masaki Y, Kurose N, Yamamoto M, Takahashi H, Saeki T, Azumi A, et al. Cutoff values of serum IgG4 and histopathological IgG4. *Int J Rheumatol* 2012. <http://dx.doi.org/10.1002/hep.26977>.
- [6] Deshpande V, Zen Y, Chan JK, et al. Consensus statement on the pathology of IgG4-related disease. *Mod Pathol* 25; 9: 1181–92.
- [7] Paik WH, Ryu JK, Park JM, Song BJ, Park JK, Kim YT, et al. Clinical and pathological differences between serum immunoglobulin G4-positive and-negative type 1 autoimmune pancreatitis. *World J Gastroenterol* 2013;19:4031–8.
- [8] Khosroshahi A, Cheryk LA, Carruthers MN, Edwards JA, Bloch DB, Stone JH. Spurious low serum IgG4 concentrations caused by the prozone phenomenon in patients with IgG4-related disease. *Arthritis Rheum* 2014;66:213–7.
- [9] McLean-Tooke A, O'Sullivan M, Easer T, Loh R. Differences between total IgG and sum of the IgG subclasses in clinical samples. *Pathology* 2013;45:675–7.
- [10] Urquhart N, Mattman A. Falsely low IgG4 levels due to antigen excess. *Clin Biochem* 2011;44(13):1177–8 [Abstract].