tvst

Article

Risks and Challenges in Interpreting Simultaneous Analyses of Multiple Cytokines

Isabel B. Pfister[1], Souska Zandi[1,2], Christin Gerhardt[1], Jan Spindler[1], Natalie Reichen[1], and Justus G. Garweg[1,2]

[1] Swiss Eye Institute, Rotkreuz, and Berner Augenklinik am Lindenhofspital, Bern, Switzerland
[2] Department of Ophthalmology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

Correspondence: Isabel B. Pfister, Berner Augenklinik am Lindenhofspital, Bremgartenstrasse 119, CH - 3012 Bern, Switzerland.
e-mail: isabel.pfister@augenklinik-bern.ch

Received: December 17, 2019
Accepted: April 21, 2020
Published: June 24, 2020

Keywords: cytokines; lowest limit of quantitation; statistical significance; multiple comparison; detection range

Citation: Pfister IB, Zandi S, Gerhardt C, Spindler J, Reichen N, Garweg JG. Risks and challenges in interpreting simultaneous analyses of multiple cytokines. Trans Vis Sci Tech. 2020;9(7):27, https://doi.org/10.1167/tvst.9.7.27

Purpose: To determine the inherent risks of handling results below the lowest detectable value in the analysis of multiple cytokines in the aqueous humor of patients with retinal diseases by comparing possible statistical strategies to lower the risk of mis interpretation or over interpretation of results. Furthermore, in analyzing multiple cytokines simultaneously, the challenge of multiple comparison arises.

Methods: The analyses were based on parallel testing of 43 cytokines in 58 aqueous humor samples from patients with macular hole or epiretinal membrane. Substitution of values below the detection limit with 0.1 ×, 0.5 ×, or 1.0× of the lowest level of quantitation was compared with handling as missing value. The impact of correction for multiple comparisons was assessed using the Holm correction.

Results: When comparing macular hole with epiretinal membrane, not substituting the missing data revealed a difference ($P < 0.05$) for five compared with wight cytokines after their substitution, indicating an increased risk for under-estimating group differences (type II error). Correcting for multiple comparisons revealed a relevant risk of over estimating group differences (type I error).

Conclusions: Physiologic cytokine concentrations in ocular fluids typically range at or below the lowest level of quantitation. Handling of results below this cutoff as missing leads to increased type II errors. Not correcting for multiple comparisons increases the risk of a type I error. Taken together, both harbor a systematic inherent risk of misinterpretation of the results.

Translational Relevance: Ignoring the inherent risks of data misinterpretation in analyses of ocular fluid samples may result in mis leading conclusions regarding their biological relevance.

Introduction

An analysis of intraocular fluids means assessing concentrations of targets, be it proteins, antibodies, cytokines, or drugs, at or below the lower end of sensitivity of commercially available test systems. The concentration of antibodies and albumin in intraocular fluids, for example, ranges between 200- and 1000-fold lower than in serum.[1,2] This condition interferes with the test reliability, because the reading out of results based on standard curves in this lower range is not very reliable or reproducible. Manufacturers of analytical devices do not advocate using data below the lower threshold of a test system, therefore these results require special consideration. Some authors propose to discard these data or suggest the out of range (OOR) data be set to zero.[3–5] Other investigators have reported substitution of concentrations below the detection limit with 0.5 times of the lowest point of the calibration curve.[6]

While evaluating our own test results generated by a multiplex beads system in cytokine/chemokine research of intraocular fluids,[7–12] we experienced several challenges with data handling. Because many researchers in this field likely have similar experiences, our goal was to systematically assess different statistical

methods and outcomes to determine the best approach to handle this type of data.

A detection limit of 1 pg is recommended for the majority of detectable cytokines by the manufacturers of sensitive immunobeads systems (e.g., Bio-Rad multiplex immunoassays; Bio-Rad, Hercules, CA). This raises the question of how to statistically handle OOR values in ocular and other samples, which may apply to a considerable proportion of samples quantified in cytokine research (In our data we rarely had the case of single data points laying above the standard curve range. If a single sample shows unexplained high values, a repeat parallel double testing should be considered. The solution to systematically high target values in a series would be a higher dilution of samples prior to submission to the test.).

Multiplex assays allow for the quantification of multiple targets in parallel in a small volume and within a single test run. The inherent problem with multiple comparisons is an increased risk for a type I error, which results in an overinterpretation of the results. Discussions and possible solutions to this problem have already been discussed since the early 1960s.[13] Although available, the statistical tools to control for multiple comparisons have as yet not routinely been established. Therefore, we addressed the impact of data handling and of correction for multiple comparisons based on a set of cytokine data derived from aqueous samples from patients with two similar and specific ocular diagnoses (macular hole [MH] and epiretinal membrane [ERM]) in the absence of any systemic disease or treatment.

Methods

Patients

This prospective clinical study included aqueous humor (AH) samples from eyes with symptomatic idiopathic ERM ($n = 29$) or MH ($n = 29$) undergoing vitrectomy.

Clinical data pertaining to the systemic and ophthalmologic diagnoses, the treatment strategy, and the duration of the ocular symptoms were documented. Patients with systemic comorbidities, that is, with diabetes mellitus, a history of intraocular surgery (other than uneventful phacoemulsification for senile cataract, performed minimally 6 months before vitrectomy), known rheumatic and autoimmune diseases, systemic treatments involving corticosteroids or immunomodulatory drugs, ocular trauma, vitreal hemorrhaging, uveitis, glaucoma, or any concomitant retinal pathology, were excluded. Also excluded were patients who had undergone intraocular surgery or treatment within 6 months before sample collection. Patients with prior surgery, such as cataract surgery, were only included in cases of uneventful surgery. No patient had a posterior capsular tear or violation. The results of the corresponding vitreous analyses have been published.[7] Surgery was performed at the Berner Augenklinik am Lindenhofspital in Bern after informed, written consent of the individuals had been obtained for vitrectomy and ocular fluid sampling. This study is fully compliant with the tenets of the Declaration of Helsinki and was approved by the Institutional Ethics Board of the University of Bern, Switzerland (registration number: 152/08).

Undiluted AH samples (approximately 100 µL) were collected at the beginning of surgery and immediately stored at –80°C until analysis. All samples were analyzed in parallel using a multiplex beads system (Bio-Plex 100 array reader with Bio-Plex Manager software version 6.1; Bio-Rad; human chemokine assay with 40 cytokines: P171-AK99MR2 and TGF-β 3-plex assay: 171W4001M). With this highly sensitive technique, multiple analytes can be detected in parallel using a single small volume sample. For the purpose of this study, we quantified the concentrations of 43 cytokines in each aqueous sample (Tables 1 and 2). All analytic procedures were performed following the manufacturer's instructions. In short, magnetic microspheres, tagged with a fluorescent label, were coupled to specific capture antibodies and mixed with samples containing unknown cytokine quantities before introducing biotinylated detection antibodies and streptavidin R-phycoerythrin. The mixture was then analyzed by flow cytometry. The instrument's two lasers identify the microsphere type and quantify the amount of bound antigen. On each test plate we ran a duplicate concentration standard in parallel for each cytokine. The measurements were performed in a blinded manner by a laboratory technician who was experienced in the execution of this technique.

Statistical Workup of Data

The concentration of cytokines below the standard curve of fit were registered as missing values and thus excluded by the automated software provided with the reader. Because the handling of these data as missing values constitutes a loss of information, we compared different options for handling of OOR values below the lower limit of quantification (LLOQ) provided by the manufacturer by either (a) accounting them as missing, (b) by substituting the OOR data with the lower limit of detection (LOD) specific in our sample, or (c) substituting them with different fractions of the LLOQ, that

Table 1. Descriptive Results for Cytokine Concentrations in the AH of 29 Eyes With MH Without and After Substitution of OOR Data With Different Proportions of the LLOQ

Cytokine (pg/mL)	*n*	OOR data	With Substitution of OOR by Missing Values		With Substitution of OOR by the LOD		With Substitution by 0.1 of the LLOQ		With Substitution by 0.5 of the LLOQ		With Substitution by 1.0 of the LLOQ		Kruskal -Wallis H Test
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
CCL21	28	1	777.9	347.2	761.8	351.8	751.2	370.1	751.5	369.5	751.9	368.7	0.9992
CXCL13	10	19	0.7	0.9	0.3	0.6	0.3	0.6	0.5	0.5	0.7	0.5	**3.77E-11**
CCL27	7	22	3.4	5.2	0.9	2.8	0.9	2.8	1.3	2.7	1.7	2.6	**4.77E-13**
CXCL5	27	2	177.3	139.4	165.3	141.6	165.1	141.8	165.3	141.6	165.5	141.3	0.9888
CCL11	28	1	5.4	2.3	5.3	2.4	5.2	2.5	5.3	2.4	5.3	2.4	0.9993
CCL24	29	0	27.2	14.4	27.2	14.4	27.2	14.4	27.2	14.4	27.2	14.4	1
CCL26	24	5	4.3	2.7	3.6	2.9	3.6	3.0	3.6	2.9	3.7	2.8	0.766
CX3CL1	29	0	43.2	17.9	43.2	17.9	43.2	17.9	43.2	17.9	43.2	17.9	1
CXCL6	9	20	1.3	1.2	0.6	0.8	0.5	0.9	0.7	0.8	1.0	0.7	**6.49E-08**
GM-CSF	29	0	145.9	97.7	145.9	97.7	145.9	97.7	145.9	97.7	145.9	97.7	1
CXCL1	27	2	43.9	22.0	41.1	23.5	40.9	24.0	41.0	23.8	41.1	23.7	0.989
CXCL2	10	19	8.1	2.3	6.2	1.9	3.1	3.9	4.3	3.1	5.8	2.1	**4.15E-06**
CCL1	24	5	12.9	8.6	11.6	8.3	10.7	9.2	10.8	9.1	11.0	8.9	0.76
IFN-γ	3	26	5.2	2.7	2.4	1.2	0.7	1.7	1.6	1.4	2.6	1.1	**1.01E-15**
IL-1β	29	0	0.9	0.8	0.9	0.8	0.9	0.8	0.9	0.8	0.9	0.8	1
IL-2	25	4	0.6	0.7	0.5	0.6	0.5	0.6	0.6	0.6	0.6	0.6	0.5298
IL-4	3	26	2.6	1.9	1.3	0.7	0.4	0.9	0.8	0.8	1.3	0.7	**2.20E-16**
IL-6	29	0	26.1	92.6	26.1	92.6	26.1	92.6	26.1	92.6	26.1	92.6	1
IL-8/CXCL8	29	0	5.6	6.4	5.6	6.4	5.6	6.4	5.6	6.4	5.6	6.4	1
IL-10	24	5	3.8	2.9	3.4	2.9	3.2	3.0	3.3	2.9	3.4	2.8	0.77
IL-16	28	1	14.8	10.7	14.3	10.8	14.8	10.7	14.3	10.8	14.8	10.7	1.00
CXCL10	29	0	43.6	34.0	43.6	34.0	43.6	34.0	43.6	34.0	43.6	34.0	1
CXCL11	25	4	1.1	0.8	1.0	0.8	0.9	0.8	0.9	0.8	0.9	0.8	0.88
CCL2	29	0	452.6	236.9	452.6	236.9	452.6	236.9	452.6	236.9	452.6	236.9	1
CCL8	29	0	2.8	1.8	2.8	1.8	2.8	1.8	2.8	1.8	2.8	1.8	1
CCL7	1	28	5.8		4.6	0.2	0.4	1.0	1.1	0.9	2.0	0.7	**2.20E-16**
CCL13	29	0	1.3	0.6	1.3	0.6	1.3	0.6	1.3	0.6	1.3	0.6	1
CCL22	23	6	9.7	4.7	8.4	4.9	7.7	5.7	7.7	5.6	7.8	5.5	0.6278
MIF	29	0	38758.7	19067.5	38758.7	19067.5	38758.7	19067.5	38758.7	19067.5	38758.7	19067.5	1
MIG/ CXCL9	29	0	14.8	18.2	14.8	18.2	14.8	18.2	14.8	18.2	14.8	18.2	1
CCL3	27	2	1.0	0.4	0.9	0.5	0.9	0.5	0.9	0.5	0.9	0.5	0.99
CCL15	29	0	516.3	532.7	516.3	532.7	516.3	532.7	516.3	532.7	516.3	532.7	1
CCL20	29	0	3.4	3.1	3.4	3.1	3.4	3.1	3.4	3.1	3.4	3.1	1
CCL19	13	16	7.4	4.9	3.4	4.9	3.5	4.8	4.2	4.4	5.0	3.9	0.001232
CCL23	21	8	16.3	9.3	12.0	10.6	11.8	10.7	11.9	10.6	12.1	10.5	0.3512
CXCL16	29	0	474.2	182.8	474.2	182.8	474.2	182.8	474.2	182.8	474.2	182.8	1
CXCL12	27	2	109.9	68.0	103.6	69.7	102.4	71.3	102.6	71.0	102.9	70.6	0.99
CCL17	2	27	3.9	2.2	0.3	1.1	0.4	1.0	1.1	0.9	1.9	0.7	**2.20E-16**
CCL25	29	0	106.0	63.1	106.0	63.1	106.0	63.1	106.0	63.1	106.0	63.1	1
TNF-α	29	0	5.6	4.5	5.6	4.5	5.6	4.5	5.6	4.5	5.6	4.5	1
TGF-β1	26	3	225.9	137.8	203.5	146.4	202.6	147.8	202.6	147.7	202.7	147.6	0.95
TGF-β2	29	0	3337.8	1435.4	3337.8	1435.4	3337.8	1435.4	3337.8	1435.4	3337.8	1435.4	1
TGF-β3	5	24	16.9	18.7	3.4	9.4	3.1	9.5	4.1	9.2	5.2	8.9	**6.93E-12**

LLOQ: Lower limit of quantitation; LOD: lower of detection; SD: Standard deviation. Significant *p*-values are marked in bold.

Table 2. Descriptive Results for Cytokine Concentrations in the AH of Eyes With an ERM Without and After Substitution of OOR Data With Different Proportions of the LLOQ

Cytokine		OOR	With Substitution of OOR by Missing Values		With Substitution of OOR by the LOD		MH With Substitution by 0.1 of the LLOQ		MH With Substitution by 0.5 of the LLOQ		MH With Substitution by 1.0 of the LLOQ		Kruskal -Wallis
Cytokine	*n*	data	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	H Test
CCL21	29	0	1140.0	662.0	1140.0	662.0	1140.0	662.0	1140.0	662.0	1140.0	662.0	1.0
CXCL13	17	12	0.4	0.3	0.3	0.3	0.3	0.3	0.4	0.3	0.5	0.3	**9.98E-05**
CCL27	5	24	1.2	1.6	0.2	0.8	0.3	0.7	0.7	0.7	1.2	0.6	**7.07E-16**
CXCL5	29	0	2444.4	3566.4	2444.4	3566.4	2444.4	3566.4	2444.4	3566.4	2444.4	3566.4	1.0
CCL11	29	0	8.8	3.1	8.8	3.1	8.8	3.1	8.8	3.1	8.8	3.1	1.0
CCL24	29	0	30.9	14.2	30.9	14.2	30.9	14.2	30.9	14.2	30.9	14.2	1.0
CCL26	20	9	3.3	2.0	2.4	2.1	2.3	2.2	2.4	2.1	2.5	2.0	0.2569
CX3CL1	29	0	39.9	19.1	39.9	19.1	39.9	19.1	39.9	19.1	39.9	19.1	1.0
CXCL6	10	19	1.7	1.1	0.8	0.9	0.6	1.0	0.8	0.9	1.1	0.7	**6.43E-06**
GM-CSF	29	0	90.4	35.3	90.4	35.3	90.4	35.3	90.4	35.3	90.4	35.3	1.0
CXCL1	27	2	47.4	18.2	44.4	20.8	44.2	21.3	44.2	21.1	44.3	20.9	0.99
CXCL2	5	24	11.8	1.4	6.3	2.6	2.4	4.4	3.9	3.7	5.8	2.8	**6.93E-12**
CCL1	3	26	12.1	5.2	6.0	2.5	1.4	4.0	2.1	3.7	2.9	3.5	**1.01E-15**
IFN-γ	6	23	5.2	3.5	2.7	2.0	1.3	2.5	2.0	2.2	2.9	1.9	**3.31E-12**
IL-1β	28	1	0.9	0.7	0.9	0.7	0.9	0.7	0.9	0.7	0.9	0.7	1.00
IL-2	24	5	0.9	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.8	0.6	0.54
IL-4	20	9	31.1	29.1	21.8	27.8	21.5	28.0	21.6	27.9	21.8	27.8	0.23
IL-6	29	0	9.0	10.8	9.0	10.8	9.0	10.8	9.0	10.8	9.0	10.8	1.0
IL-8 /CXCL8	29	0	9.1	7.1	9.1	7.1	9.1	7.1	9.1	7.1	9.1	7.1	1.0
IL-10	29	0	5.5	1.8	5.5	1.8	5.5	1.8	5.5	1.8	5.5	1.8	1.0
IL-16	28	1	16.9	9.0	16.4	9.3	16.9	9.0	16.3	9.3	16.9	9.0	1.00
CXCL10	29	0	42.9	24.7	42.9	24.7	42.9	24.7	42.9	24.7	42.9	24.7	1.0
CXCL11	29	0	1.0	0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0
CCL2	29	0	452.3	262.6	452.3	262.6	452.3	262.6	452.3	262.6	452.3	262.6	1.0
CCL8	29	0	2.9	1.6	2.9	1.6	2.9	1.6	2.9	1.6	2.9	1.6	1.0
CCL7	4	25	8.9	2.9	5.2	1.8	1.4	3.2	2.0	3.0	2.9	2.6	**2.74E-14**
CCL13	29	0	2.1	1.3	2.1	1.3	2.1	1.3	2.1	1.3	2.1	1.3	1.0
CCL22	24	5	13.2	6.6	11.5	7.1	10.9	7.8	11.0	7.7	11.1	7.6	0.77
MIF	29	0	14260.8	25156.5	14260.8	25156.5	14260.8	25156.5	14260.8	25156.5	14260.8	25156.5	1.0
MIG/ CXCL9	28	1	12.7	7.8	12.4	7.9	12.3	8.0	12.3	8.0	12.3	8.0	1.00
CCL3	26	3	1.2	0.5	1.1	0.6	1.1	0.6	1.1	0.6	1.1	0.6	0.95
CCL15	29	0	497.4	326.7	497.4	326.7	497.4	326.7	497.4	326.7	497.4	326.7	1.0
CCL20	29	0	3.8	3.3	3.8	3.3	3.8	3.3	3.8	3.3	3.8	3.3	1.0
CCL19	26	3	9.3	4.6	8.3	5.1	8.3	5.1	8.5	4.9	8.6	4.7	0.95
CCL23	27	2	14.7	7.5	13.7	8.1	13.7	8.2	13.7	8.1	13.7	8.1	0.99
CXCL16	29	0	588.4	235.6	588.4	235.6	588.4	235.6	588.4	235.6	588.4	235.6	1.0
CXCL12	29	0	107.5	59.5	107.5	59.5	107.5	59.5	107.5	59.5	107.5	59.5	1.0
CCL17	1	28	0.1		0.1	0.0	0.2	0.0	0.8	0.1	1.6	0.3	**2.20E-16**
CCL25	29	0	56.6	55.1	56.6	55.1	56.6	55.1	56.6	55.1	56.6	55.1	1.0
TNF-α	29	0	6.2	2.8	6.2	2.8	6.2	2.8	6.2	2.8	6.2	2.8	1.0
TGF-β1	21	8	157.7	174.6	116.8	162.2	114.3	164.1	114.4	163.9	114.7	163.7	0.34
TGF-β2	29	0	2910.6	1033.6	2910.6	1033.6	2910.6	1033.6	2910.6	1033.6	2910.6	1033.6	1.0
TGF-β3	19	10	14.1	23.5	9.4	20.0	9.3	20.0	9.7	19.8	10.2	19.7	0.10

LLOQ: Lower limit of quantitation; LOD: lower of detection; SD: Standard deviation. Significant *p*-values are marked in bold.

is, a factor of 0.1, 0.5, or 1.0 (fractions) of the LLOQ. Numerical and statistical outcomes were compared in relation to the handling of results in fractions (LOD; 0, 0.1, 0.5, or 1.0-fold the LLOQ).

Because most of the data were not normally distributed, a series of nonparametric Kruskal-Wallis H tests were applied to compare the different substitutions of the data below the detection limit (OOR< data). Mann–Whitney *U* tests were applied for each cytokine to explore the difference between MH and ERM. Statistical evaluation was performed using the R statistical package psych (version 3.2.4; R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2016).

Because multiple comparisons increase the risk of introducing a type I error (declaring a result statistically significant when it is not), we adjusted the significance level to prevent falsely claiming statistical significance.[14] The classic approach to the multiple comparison problem is to control the family-wise error rate.[15] To escape this problem, the critical *P* level for significance, or alpha, is frequently arbitrarily set more restrictively (i.e., $P < 0.01$). One relatively simple approach is the application of Bonferroni's correction.[16] This correction is a rather conservative estimate that in turn leads to a higher risk for type II errors (not declaring a result as statistically significant, when it is in fact significant).[14] To control for type I errors, but without driving up the risk of type II errors, the Holm correction was established as a more qualified approach[17,18] and applied here for correction of the statistical outcomes.

Holm's correction is a sequentially rejective Bonferroni test that progressively adapts the threshold for rejecting the null hypothesis. As a first step, all *P* values are sorted of smallest to largest, with *k* indicating the number of hypotheses. In a second step, the lowest *P* value is compared with α/k. If the *P* value is lower, the null hypothesis is rejected and the significance of a result is confirmed. The same procedure is applied to the remaining $k - 1$ hypotheses where the threshold of significance is set at $\alpha/(k - 1)$. This procedure is repeated sequentially for each *P* value, until the selected *P* value is higher than the sequential threshold.

Results

Patients of both MH and ERM groups were comparable regarding age at the time of inclusion (70.0 ± 8.7 and 67.4 ± 8.0 years, respectively; $P = 0.17$), as well as sex (80% females in the ERM group compared with 55% females in the MH group; $P = 0.052$).

Handling of OOR Data

The AH samples in MH provided results below the detection threshold for 25 out of the 43 cytokines (OOR<: CCL21, CXCL13, CCL27, CXCL5, CCL11, CCL26, CXCL6, CXCL1, CXCL2, CCL1, interferon [IFN]-γ, IL-2, IL-4, IL-10, IL-16, CXCL11, CCL7, CCL22, CCL3, CCL19, CCL23, CXCL12, CCL17, transforming growth factor [TGF]-β1, and TGF-β3; Table 1). For nine cytokines (20.9%), more than two-thirds of the samples (CXCL13, CCL27, CXCL6, CXCL2, IFN-γ, IL-4, CCL7, CCL17, and TGF-β3) ranged below the lower LOD (Table 1).

In samples from eyes with ERM, 21 of 43 cytokines presented results below the detection threshold (OOR<: CXCL13, CCL27, CCL26, CXCL6, CXCL1, CXCL2, CCL1, IFN-γ, IL-1β, IL-2, IL-4, IL-16, CCL7, CCL22, MIG/CXCL9, CCL3, CCL19, CCL23, CCL17, TGF-β1 and TGF-β3; Table 2). Again, in seven of these (16.3%) cytokines, more than two-thirds of all samples tested (CCL27, CXCL6, CXCL2, CCL1, IFN- γ, CCL7, and CCL17) ranged below the lower LOD (Table 2).

Comparing MH and ERF samples with and without the substitution of the OOR data revealed a relevant difference in the mean values and standard deviations for all cytokines in which more than two-thirds of the results ranged below the detection limit (Tables 1 and 2). It remains to be discussed, whether a statistical comparison in this situation is meaningful at all.

If the OOR data were handled as missing values, a significant difference between MH and ERM was found for 14 cytokines. When the OOR data were substituted with any of the fractions of the LLOQ, a significant difference between the two groups was also observed for 14 cytokines. The substitution with the LOD or with a factor of 0.1 to 1.0 of the LLOQ led to similar results (Table 3).

Correcting for Multiple Comparisons

When the OOR data were not substituted, after applying the Holm correction, the difference remained significant for 5 of the 14 cytokines (Table 3). The difference in the number of significances in the comparison of ERM and MH (five vs 14) was significant ($P = 0.019$). After substitution of OOR data, and applying the Holm correction, the difference remained significant for 8 of the 14 cytokines (Table 3). However, the difference in the number of significances in the comparison of ERM and MH (eight vs. 14) did not remain significant ($P = 0.14$). The different substitutions

Table 3. Statistical Outcomes After Comparison of the Cytokine Concentrations Within the AH in MH and ERM With Substitution of OOR Data With Half of the LLOQ

	Comparison Between MH and ERM With No Substitution of OOR		Comparison Between MH and ERM With Substitution of OOR by LOD		Comparison Between MH and ERM With Substitution of OOR by o.5 the LLOQ	
Cytokine	Mann-Whitney U test	Holm's correction	Mann-Whitney U test	Holm's correction	Mann-Whitney U test	Holm's correction
CL21	**0.008444**		**0.005246**		**0.005247**	
CXCL13	0.6151		0.2729		0.8327	
CCL27	0.9352		0.3572		0.9912	
CXCL5	0.3889		0.2102		0.216	
CCL11	**4.99E-05**	**Sig.**	**2.95E-05**	**Sig.**	**2.95E-05**	**Sig.**
CCL24	**0.04914**		**0.04914**		**0.04914**	
CCL26	0.3395		0.09494		0.07539	
CX3CL1	0.3234		0.3234		0.3234	
CXCL6	0.4065		0.3999		0.2436	
GM-CSF	**0.0001522**	**Sig.**	**0.0001522**	**Sig.**	**0.0001522**	**Sig.**
CXCL1	0.4208		0.4502		0.4693	
CXCL2	**0.01102**		0.4009		0.2884	
CCL1	0.906		**7.22E-07**	**Sig.**	**2.89E-07**	**Sig.**
IFN-γ	1		0.6306		0.283	
IL-1β	0.5762		0.7439		0.7263	
IL-2	**0.03306**		0.07848		**0.04296**	
IL-4	**0.02215**		**8.60E-07**	**Sig.**	**1.75E-06**	**Sig.**
IL-6	0.8887		0.8887		0.8887	
IL-8 /CXCL8	**0.01431**		**0.01431**		**0.01431**	
IL-10	**0.0006389**	**Sig.**	**4.12E-05**	**Sig.**	**4.12E-05**	**Sig.**
IL-16	0.06857		0.08816		0.08396	
CXCL10	0.7032		0.7032		0.7032	
CXCL11	0.8554		0.3009		0.2866	
CCL2	0.6242		0.6242		0.6242	
CCL8	0.6575		0.6575		0.6575	
CCL7	0.4682		0.2802		0.1551	
CCL13	**0.0192**		**0.0192**		**0.0192**	
CCL22	0.07357		0.08864		0.1241	
MIF	**3.82E-07**	**Sig.**	**3.82E-07**	**Sig.**	**3.82E-07**	**Sig.**
MIG/ CXCL9	0.5441		0.7089		0.7147	
CCL3	**0.01683**		0.06466		0.0625	
CCL15	0.4601		0.4601		0.4601	
CCL20	0.5702		0.5702		0.5702	
CCL19	0.2327		**0.0002582**	**Sig.**	**0.0002039**	**Sig.**
CCL23	0.4861		0.395		0.4041	
CXCL16	0.05098		0.05098		0.05098	
CXCL12	0.9804		0.6688		0.6689	
CCL17	0.2207		0.1537		0.08507	
CCL25	**9.16E-05**	**Sig.**	**9.16E-05**	**Sig.**	**9.16E-05**	**Sig.**
TNF-α	0.09276		0.09276		0.09276	
TGF-β1	**0.03058**		**0.00625**		**0.006766**	
TGF-β2	0.2662		0.2662		0.2662	
TGF-β3	0.3009		**0.001882**		**0.00419**	
No. of significant results with NO correction	14		14		14	
WITH Holm correction		5		8		8

LLOQ: Lower limit of quantitation. LOD: Lower of detection. Significant *p*-values are marked in bold.

(0.1×, 0.5×, and 1.0×) of values below the LLOQ and the LOD did not impact this outcome.

Discussion

Pro- and anti-inflammatory cytokines are partly upregulated in the AH of patients with ERM, which is in line with our previous findings of 2016 regarding the vitreous,[7] although findings from AH cannot be expected to correlate with those from the vitreous fluid.[19]

Of the cytokines that showed a significant difference between the two pathologic groups, CCL11, GM-CFS, MIF, CCL25, CCL1, and CCL19 belong to the proinflammatory cytokines, partially with chemoattractant properties. Of these, CCL11 and CCL19 where upregulated for patients with ERM. The two anti-inflammatory cytokines, IL-10 and IL-4, were higher for patients with ERM compared with MH. As mentioned in our article from 2016,[7] because the levels of given cytokines that are necessary for intercellular communication and for the driving of relevant changes in this milieu are still not well-known, we are unable to assess the biological relevance of the measured differences between the two pathologic groups.

In our study, the concentration of several cytokines was below the standard curve. However, these data still may provide important information, namely, that the concentrations of these cytokines are low in the corresponding samples, which is relevant for the interpretation of the mean. The exclusion of these results from the calculations necessarily results in an upwards shift of averages. Therefore, missing values can lead to misinterpretation of statistical outcomes which can be overcome by their substitution with the LOD or a fraction (0.1–1.0 times) of the predefined LLOQ.

However, it remains to be determined whether or when a statistical comparison in this situation is still meaningful. We postulate, but cannot provide firm evidence, that if more than 50% of data are below the LLOQ, then the corresponding biomarker, in our series cytokine, does not likely play a relevant role for the bioregulation of the assessed environment. Because a downregulation of the corresponding biomarker may have taken place, we think that completely ignoring biomarkers based on a high number of OOR data is more dangerous than replacing them by 0.1- to 1.0-fold the LLOQ or the LOD.

To corroborate the raw data, it may thus be advisable first to replace the missing values for results below the LLOQ by a multiplicity of 0.1 to 1.0 of the LLOQ or the LOD before applying a correction for multiple comparison. Using this strategy reduced the number of comparisons achieving significance from 33% (14 of 43) using the default setting of not weighing OOR values and using a significance level to a P value of less than 0.05 as in the automated statistical software, to 19% (8 of 43) in our model.

Assessing a single cytokine may reveal statistical significance, but it is difficult to interpret its relevance in a biological process where the pathophysiology of a disease is not fully understood. The ease of comparing multiple analytes in a single set of samples, in contrast, may produce a lot of complex and interrelated data regarding the cytokine environment and its changes in pathologic states, which harbors an inherent risk of misinterpretation. The analysis of a single or few cytokines may be of interest to analyze specific biological processes, like specific immune responses. The evaluation of multiple cytokines in parallel, in contrast, may provide important insights into the local cytokine environment in unknown or less specifically regulated processes.

Data Below the Detection Limit (OOR<)

Handling OOR data as missing values may formally be correct, but will not describe the biological regulation in an environment such as the eye, which is known to have biomarker concentrations at the lower end, at least in healthy states. Therefore, minute changes to this equilibrium may very well be meaningful. Excluding these data would prevent detecting a possible downregulation of distinct markers and thus may lead to a relevant bias and a potential underestimation of group differences (Table 3). As mentioned, which fraction of the LLOQ or the LOD were substituted had minimal influence on the results. Substitution of the OOR data with the LOD or different proportions of the LLOQ (0.1, 0.5, and 1.0 of the LLOQ) revealed a difference in means between MH and ERF for eight cytokines, whereas no handling of the OOR data revealed a difference in only five cytokines, which would have failed to identify significance for 3 of 43 cytokines (7%). The consequent underestimation of group differences increases the risk for a type II error.

Correcting for Multiple Comparisons

If large numbers of independent statistical tests are performed on the same sample, the probability of obtaining significant results will increase. That is, some P values will be less than 0.05 just by chance. This phenomenon results from choosing a significance level of 0.05 which, by definition, results on average in 1 out of 20 comparisons being declared significant by

chance alone, even if there is in fact no real difference between the groups. In other words, performing multiple comparisons increases the risk of a type I error. The necessity to correct for multiple comparisons has generally been accepted in the scientific community and can be done by different methods like the Bonferroni correction.[17] Less conservative methods are Holm's correction,[18] Hochberg (1988),[20] Hommel (1988),[21] Benjamini and Hochberg (1995),[22] and Benjamini and Yekutieli (2001).[23] The first four methods control for the family-wise error rate (Family-wise error rate is the probability of one or more false rejections. The term "family" refers to the collection of hypotheses H_1,..... H_s, which are being considered for joint testing[17]. Type I errors can be defined as family-wise error rates[23]). Hochberg's and Hommel's methods are best implemented when the hypothesis tests are independent or when they are non-negatively associated.[21] Hommel's method is more robust to decrease a type I error than Hochberg's, but the difference is usually small and the Hochberg P values are easier to compute. The methods of Benjamini, Hochberg, and Yekutieli control the false discovery rate,[22,23,26] that is, the expected proportion of false positives among the rejected hypotheses. The false discovery rate is a less stringent condition than the family-wise error rate, so these methods are more powerful than the others.[17]

The Bonferroni adjustment or correction is one of the most often used methods; it is the easiest to compute and uses a statistical significance level based on the level for a single hypothesis divided by the number of hypotheses tested. Thus, instead of setting the level of significance at 0.05, the P value for significance is divided by the number of tests performed. In our case, we tested the difference in 43 cytokines between two groups: this led to a P level for significance of $0.05/43 = 0.0012$ (i.e., only results with a P value of less than 0.0012 are accepted as significant). However, this correction is a very conservative solution, which in turns leads to a higher risk for type II errors.[14] The Holm correction offers a good solution[17,18] for controlling type I errors, but at the same time without drastically driving up type II errors, and is still easy to compute. Holm's correction, also called the sequentially rejective Bonferroni test, progressively adapts the threshold for rejecting the null hypotheses.

For our data, we chose to apply the Holm correction because it is not as restrictive as the Bonferroni correction, but still simple enough that it can be easily applied. The application of a correction for multiple comparisons reduced the number of statistically significant differences by 50%: instead of 14 cytokines with a significant difference between the two groups, we found 5 ($P = 0.019$) or 8 ($P = 0.14$), depending on whether we substituted the OOR data or not. Much time and effort for future research could be saved by not pursuing nonexisting differences.

In conclusion, we were able to demonstrate a relevant impact of handling OOR data to describe a biological environment, and the importance of correction for multiple comparisons of samples with expected low concentrations to understand their potential role in a biological process. We strongly believe that any biological process involves multiple interactions of cytokines. Understanding and interpreting a pathophysiologic process might thus be based on a broad (multiplex) analysis of the local cytokine environment instead of single cytokines, but has to provide measures to escape the inherent risks of multiple comparisons.

Acknowledgments

Disclosure: **I.B. Pfister,** None; **S. Zandi**, None; **C. Gerhardt**, None; **J. Spindler**, None; **N. Reichen**, None; **J.G. Garweg**, Novartis (C), Bayer (C), Alcon (C), AbbVie (C)

References

1. Garweg JG, Boehnke M. The antibody response in experimental ocular toxoplasmosis. *Graefes Arch Clin Exp Ophthalmol*. 2006;244:1668–1679.
2. Garweg JG, Ventura AC, Halberstadt M, et al. Specific antibody levels in the aqueous humor and serum of two distinct populations of patients with ocular toxoplasmosis. *Int J Med Microbiol*. 2005;295:287–295.
3. Cheung CMG, Vania M, Ang M, Chee SP, Li J. Comparison of aqueous humor cytokine and chemokine levels in diabetic patients with and without retinopathy. *Mol Vis*. 2012;18:830–837.
4. Takeuchi M, Sato T, Tanaka A. et al. Elevated levels of cytokines associated with Th2 and Th17 cells in vitreous fluid of proliferative diabetic retinopathy patients. *PLoS One*. 2015;10:e0137358.
5. Dong N, Xu B, Chu L, Tang X. Study of 27 aqueous humor cytokines in type 2 diabetic patients with or without macular edema. *PLoS One*. 2015;10:e0125329.
6. Valentincic NV, de Groot-Mijnes JDF, Kraut A, Korosec P, Hawlina M, Rothova A. Intraocular and serum cytokine profiles in patients with intermediate uveitis. *Mol Vis*. 2011;17:2003–2010.
7. Zandi S, Tappeiner C, Pfister IB, Despont A, Rieben R, Garweg JG. Vitreal cytokine profile

differences between eyes with epiretinal membranes or macular holes. *Invest Ophthalmol Vis Sci*. 2016;57:6320–6326.
8. Zandi S, Pfister IB, Traine PG, et al. Biomarkers for PVR in rhegmatogenous retinal detachment. *PLoS One*. 2019;14:e0214674.
9. Garweg JG, Zandi S, Pfister I, Rieben R, Skowronska M, Tappeiner C. Cytokine profiles of phakic and pseudophakic eyes with primary retinal detachment. *Acta Ophthalmol*. 2019;97:e580–e588.
10. Spindler J, Zandi S, Pfister IB, Gerhardt C, Garweg JG. Cytokine profiles in the aqueous humor and serum of patients with dry and treated wet age-related macular degeneration. *PLoS One*. 2018;13:e0203337.
11. Garweg JG, Zandi S, Pfister IB, Skowronska M, Gerhardt C. Comparison of cytokine profiles in the aqueous humor of eyes with pseudoexfoliation syndrome and glaucoma. *PLoS One*. 2017;12:e0182571.
12. Garweg JG, Zandi S, Gerhardt C, Pfister IB. Isoforms of TGF-β in the aqueous humor of patients with pseudoexfoliation syndrome and a possible association with the long-term stability of the capsular bag after cataract surgery. *Graefes Arch Clin Exp Ophthalmol*. 2017;255:1763–1769.
13. Dunn OJ. Multiple comparison among means. *J Am Stat Assoc*. 1961;56:52–64.
14. Steiner DL, Norman GR. Correction for multiple testing. Is there a resolution? *Chest*. 2011;40:16–18.
15. Bender R, Lange S. Adjusting for multiple testing – when and how? *J Clin Epidemiol*. 2001;54:343–349.
16. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. 1936;8:3–62.
17. Lehmann EL, Romano JP. Generalizations of the familywise error rate. *Ann Stat*. 2005;33:1138–1154.
18. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;6:65–70.
19. Ecker SM, Hines JC, Pfahler SM, Glaser BM. Aqueous cytokines and growth factor levels do not reliably reflect those levels found in the vitreous. Mol Vis. 2011;17:2856–2863.
20. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988;75:800–802.
21. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*. 1988;75:383–386.
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289–300.
23. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 2001;29:1165–1188.
24. Blakesley RE, Mazumdar S, Dew MA, et al. Comparisons of methods for multiple hypothesis testing in neuropsychological research. *Neuropsychology*. 2009;23:255–264.
25. Sarkar S. Some probability inequalities for ordered MTP2 random variables: a proof of Simes conjecture. *Ann Stat*. 1998;26:494–504.
26. Storey J. A direct approach to false discovery rates. *J R Stat Soc B*. 2002;64:479–498.