Variability in the Laboratory Measurement of Cytokines

A Longitudinal Summary of a College of American Pathologists Proficiency Testing Survey

Vijaya Knight, MD, PhD; Thomas Long, MPH; Qing H. Meng, PhD; Michael A. Linden, MD, PhD; Daniel D. Rhoads, MD

• ***Context.***—**The measurement of cytokines in clinical laboratories is becoming an increasingly routine part of immune monitoring when administering biologic and cell-based immunotherapies and also for clinical assessment of inflammatory conditions. While a number of commercial assays and platforms are available for cytokine measurement, there is currently little standardization among these analytical methods.**

Objective.—**To characterize the variability and comparability among cytokine testing platforms that are commonly used in clinical laboratories.**

Design.—**We analyzed data for 4 cytokines (interleukin [IL]-1, IL-6, IL-8, and tumor necrosis factor-alpha [TNF-α]) from 6 College of American Pathologists cytokine surveys administered from 2015 to 2018. Analyses interrogated variability between testing methods and variability within each laboratory across the mailings.**

Results.—**Significant variability was noted across methods with analysis of IL-1 showing the least variability and IL-6, IL-8, and TNF-α varying between methods to a greater extent. Intralab variability was also significant with TNF-α measurements again showing the greatest variability.**

Conclusions.—**This retrospective analysis of College of American Pathologists proficiency testing data for cytokine measurement is the largest method comparison to date, and this study provides a description of the variation of cytokine measurement across methods, across laboratories, and within laboratories. Serial monitoring of cytokines should preferentially be performed by the same method within the same laboratory.**

(*Arch Pathol Lab Med.* 2020;144:1230–1233; doi: 10.5858/arpa.2019-0519-CP)

Cytokines are small proteins or glycoproteins that are secreted by a variety of immune and nonimmune cells. Cytokines are responsible for a variety of pleiotropic effects, including, but not limited to, regulation of innate and adaptive immunity, communication between and among various cell types, and regulation of inflammation.[1] Serum cytokines are being measured and monitored more routinely because of their increased use as (1) therapeutics (eg, granulocyte colony-stimulating factor for treatment of neutropenia and for stem cell mobilization,[2,3] interleukin (IL)-2 for cancer immunotherapy),[4] (2) as targets for modulation of undesirable inflammation (eg, tumor necrosis factor-alpha [TNF-α][5] and IL-5),[6] (3) as biomarkers of inflammation (eg, IL-6),[7] and (4) as diagnostic markers (eg, cytokine measurement to support the diagnosis of autoinflammatory syndromes, such as Familial Mediterranean Fever).[8] Monitoring cytokine levels is used in part to detect and manage adverse effects, such as cytokine storm or cytokine release syndrome,[9,10] and cytokines are also measured during early phases of drug development to assess the potential inflammatory effects of investigational new drugs.[11] As clinicians and researchers increasingly desire to measure and monitor cytokine concentrations with a goal of drawing meaningful conclusions, it becomes increasingly important for laboratories to recognize and communicate the accuracy and reproducibility with which these analytes can be measured. This information is needed in order for laboratories to accurately interpret changes in cytokine measurements from a single person over time and to make meaningful conclusions when comparing measurements performed at different sites and in different studies.

Multiple methods are currently used to measure cytokines. Clinical laboratories measure individual cytokines separately or in panels. For example, IL-6 is widely used as a surrogate marker for inflammation, or a panel of cytokines may be used to broadly assess the overall proinflammatory

Accepted for publication March 4, 2020.
Published online May 13, 2020.
Supplemental digital content is available for this article at www.archivesofpathology.org in the October 2020 table of contents.

From the Department of Pediatrics, Section of Allergy and Immunology, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora (Knight); Department of Biostatistics, College of American Pathologists, Northfield, Illinois (Long); Laboratory Medicine, The University of Texas MD Anderson Cancer Center, Houston (Meng); Laboratory Medicine and Pathology, University of Minnesota, Minneapolis (Linden); Department of Pathology, Case Western Reserve University, Cleveland, Ohio (Rhoads).

The authors have no relevant financial interest in the products or companies described in this article.

Corresponding author: Vijaya Knight, MD, PhD, University of Colorado School of Medicine and Children's Hospital Colorado, Department of Pediatrics, Section of Allergy and Immunology, 13123 East 16th Ave, Aurora, CO 80045 (email: vijaya.knight@childrenscolorado.org).

Table 1. CTKN Survey Study Participation

Analyte	Method	CTKN Survey, N																	
		15B			16A			16B			17A			17B			18A		
		Low	Med	High	Low	Med	High	Low	Med	High	Low	Med	High	Low	Med	High	Low	Med	High
IL-1	EIA/ELISA	3	3	3	3	2	2	3	3	3	3	3	3	3	3	3	4	4	3
	Multiplex bead	4	4	5	5	5	5	4	4	4	6	6	6	4	4	4	3	3	3
	xMAP technology	6	6	6	7	7	7	7	7	7	5	5	5	5	5	5	5	5	5
IL-6	Chemiluminescence	5	5	5	4	4	4	3	3	3	3	3	2	6	5	6	4	4	3
	EIA/ELISA	13	13	13	13	12	12	13	12	11	11	10	10	9	8	8	9	6	7
	Electrochemiluminescence	4	4	4	5	5	5	3	3	3	5	5	5	6	6	6	5	5	5
	Multiplex bead	7	7	6	6	6	6	5	5	5	6	5	5	5	5	5	4	3	3
	xMAP technology	6	7	7	9	8	9	10	10	10	7	6	7	10	10	10	9	6	7
IL-8	EIA/ELISA	5	5	5	5	5	5	5	5	5	6	5	5	6	6	6	6	5	5
	Multiplex bead	3	3	3	3	3	3	2	2	2	4	4	4	2	2	2	2	2	2
	xMAP technology	4	4	4	6	6	6	5	5	5	3	3	3	5	5	5	5	5	5
TNF-α	Chemiluminescence	5	5	5	5	5	5	5	5	5	2	2	2	3	3	2	2	2	2
	EIA/ELISA	8	8	8	8	8	7	7	7	7	8	8	6	7	7	6	7	6	5
	Multiplex bead	7	7	7	5	5	5	4	4	4	6	6	6	4	4	4	3	3	3
	xMAP technology	7	7	7	10	10	10	10	10	10	8	8	8	9	9	9	10	10	10

Abbreviations: EIA, enzyme immunoassay; ELISA, enzyme-linked immunosorbent assay; IL, interleukin; Med, medium; multiplex bead, multiplex bead immunoassay; TNF-α, tumor necrosis factor-alpha; xMAP, magnetic bead–based multiplex immunoassay.

state of the individual, or to assess potential skewing toward a T-helper 1 or 2 phenotype. Measurement of a single cytokine is commonly performed using enzyme-linked immunosorbent assay or enzyme immunoassay, or chemiluminescent assays, and multiplex bead-based assays are often used to interrogate multiple cytokines using a single test.

Accuracy and precision within and between assays can be variable. Even if assays perform with good precision within an individual laboratory, the potential for interlaboratory variability is high because of variability in antibodies, calibration standards, detection reagents, detection methods, and data analysis methods.[12,13] Proficiency testing (PT) programs provide laboratories with well-defined samples designed to enable interlaboratory comparison, but the evaluation of PT data is challenging when interlaboratory variability is high, such as in the measurement of cytokines. In an effort to characterize the variability and comparability among testing platforms for 4 commonly tested cytokines (IL-1, IL-6, IL-8, and TNF-α), we analyzed data from 6 College of American Pathologists cytokine surveys collected from PT participant peer laboratories between 2015 and 2018. Analyses interrogated variability between testing methods and variability within each laboratory across the mailings.

METHODS

PT survey results reported to the College of American Pathologists in response to the cytokine 2015B, 2016A, 2016B, 2017A, 2017B, and 2018A surveys were analyzed. Reportable analytes in the surveys included interferon-gamma, IL-1 beta, IL-2, IL-6, IL-8, IL-10, TNF-α, and vascular endothelial growth factor. Each of the 6 surveys rotated the same 3 sample lots as low, medium, and high levels for each analyte. These samples were lyophilized human sera that were reconstituted by each participating laboratory. Testing methodology was reported by each participating laboratory.

Statistical analysis was performed using SAS, Inc (Cary, North Carolina). Four analytes with the greatest number of responses were analyzed in the study, including IL-1, IL-6, IL-8, and TNF-α. The 2 testing methodologies reported by participants were method and instrument. Because of sample size considerations, results were categorized into 3 to 5 technical method groups, depending upon the analyte.

Two methods were used to remove a small number of outliers as follows: (1) unreasonable values clearly outside of the distribution of results, and (2) 2 pass 3 SD outlier screen. An analysis of variation model was run to test if there were differences between the mailings for each analyte and level, because it was desired to combine data across the mailings for further statistical comparisons. Differences between methods were considered to be statistically significant if $P < .05$.

Intralab variability (coefficient of variation [CV]) was calculated for each laboratory with at least 5 mailings of results for each analyte and level. Analysis of variation was also employed to test for significant differences between the method groups for each analyte and level. Testing for intralab CV differences between the method groups could not be conducted because of the low sample sizes.

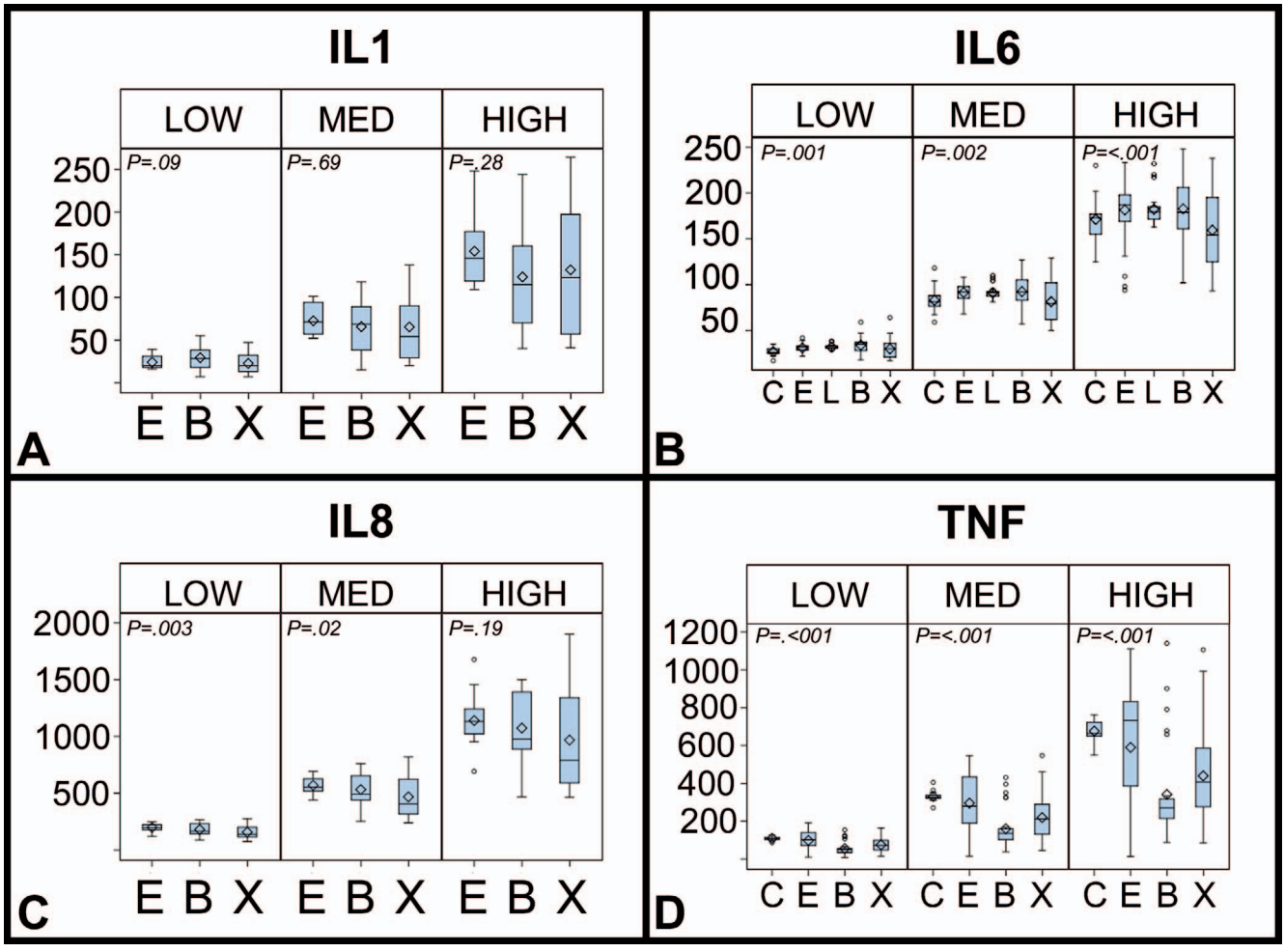
RESULTS

Study Participation

A summary of the data, including the technical method groups used for measuring each cytokine and the number of participating laboratories, is shown in Table 1. IL-6 was the most frequently analyzed analyte and enzyme immunoassay or enzyme-linked immunosorbent assay was the most commonly used method. Standard quality assurance and employment of analysis of variation indicated there was no significant difference in results across the mailings for each analyte and level, therefore suggesting the PT materials remained stable in lyophilized form throughout the duration of the analysis period. Data across the mailings were therefore aggregated for further statistical analyses.

Intermethod Comparisons

Five methods (enzyme-linked immunosorbent assay or enzyme immunoassay, chemiluminescence, electrochemiluminescence, multiplex bead immunoassay, and magnetic



Method comparison. Data for each of the 4 cytokines were grouped by method and represented as the mean (diamond symbol), median (bar), 25th and 75th percentiles, and lowest and highest values. A, Interleukin (IL)-1; B, IL-6; C, IL-8, and D, tumor necrosis factor-alpha (TNF-α). Abbreviations: E, enzyme immunoassay or enzyme-linked immunosorbent assay; B, multiplex bead immunoassay; X, xMAP technology; C, chemiluminescence; L, electrochemiluminesce.

bead–based multiplex immunoassay) were compared. Distributions of results between each method group are displayed as boxplots for each analyte (Figure). No statistically significant difference in means between method groups was observed for IL-1 (low, $P = .09$; medium, $P = .69$, and high, $P = .28$; Figure, A) and for the high level of IL-8 ($P = .19$). IL-6, IL-8, and TNF-α demonstrated statistically significant differences in the method means for each of the 3 concentrations tested, with the exception of the high level of IL-8 (Figure, B through D). The greatest variability among methods was noted for TNF-α measurement. IL-8 analyzed by the magnetic bead–based multiplex immunoassay method (Luminex; Millipore Corporation) showed consistently lower measured concentrations of IL-8 than other methods. Bead multiplex assays (other than magnetic bead–based multiplex immunoassay) measured consistently lower concentrations of TNF-α than other methods.

Intralab Variability

Because identical aliquots of PT material are sent for multiple mailings, intralab precision was determined by calculating the CV for results obtained by a single lab for an individual analyte and level. Intralab CV was calculated only when a participating lab had performed at least 5 PT surveys (Table 2). Additional data for variability (%CV) within individual laboratories participating in the surveys is presented in Supplemental Table 1 (see supplemental digital content at www.archivesofpathology.org in the October 2020 table of contents). Intralab variability was greatest for TNF-α, regardless of method and ranged from 6.7% to 102%. Intralab CVs for IL-1 measurement were less variable (4.3%–53.7%) and the least variability was noted for IL-6 and IL-8 regardless of method. However, intralab CVs for these analytes still ranged from 3.7 to 39.3, indicating that intralab variability for cytokine measurement in general tends to be significant.

DISCUSSION

This study describes the variability within and between methods and variability within an individual laboratory for measurement of IL-1, IL-6, IL-8, and TNF-α. As investigators and medical practitioners increasingly use cytokine monitoring for evaluating immune response, it is important to recognize the variability in laboratory measurement that currently exists. Our data indicate variability between methods differs for each analyte; less variability was noted

Table 2. Intralaboratory Variability by Method for Each Cytokine

			% CV		
	Variable	N	Low	Median	High
IL-1 - Method					
EIA/ELISA	IL1 - low	3	6.7	17.3	40.6
	IL1 - medium	3	4.3	11	23.2
	IL1 - high	3	6.6	14.8	28.7
Multiplex bead	IL1 - low	4	10	15.8	27.8
	IL1 - medium	4	12.7	21.2	41.6
	IL1 - high	4	7.5	18.9	26.9
xMAP technology	IL1 - low	5	16.5	21.9	53.7
	IL1 - medium	5	12.6	19.4	38.3
	IL1 - high	5	8.8	19.2	29.3
IL-6 - Method					
EIA/ELISA	IL6 - low	8	7.1	11.2	20.6
	IL6 - medium	7	4.9	7.5	11.7
	IL6 - high	7	6.1	7.7	11.1
Multiplex bead	IL6 - low	4	8.4	14.6	18.9
	IL6 - medium	3	14.5	14.9	18.1
	IL6 - high	3	9.2	10.4	18.2
xMAP TECHNOLOGY	IL6 - low	4	16.3	20	31.9
	IL6 - medium	4	13.6	18.8	25.2
	IL6 - high	4	12.6	16.9	22
IL-8 - Method					
EIA/ELISA	IL8 - low	5	5.2	11.3	14.7
	IL8 - medium	5	8.1	10.9	13.6
	IL8 - high	5	9.5	11.3	17.3
Multiplex bead	IL8 - low	2	6.3	8.8	11.3
	IL8 - medium	2	8.3	10.6	12.8
	IL8 - high	2	3.7	7.6	11.4
xMAP technology	IL8 - low	3	14.5	23.6	26.7
	IL8 - medium	3	9.7	22.9	27.8
	IL8 - high	3	6.8	21.3	39.3
TNF-α Method					
EIA/ELISA	TNF - low	5	8.9	20.7	67.8
	TNF - medium	5	6.7	24.5	87.1
	TNF - high	3	11	16.6	102
Multiplex bead	TNF - low	4	9.9	22.6	37.1
	TNF - medium	4	12	23.6	36
	TNF - high	4	10.6	22.9	39.9
xMAP technology	TNF - low	7	10.6	21.9	34.3
	TNF - medium	7	9.9	21.4	29.3
	TNF - high	7	10.2	21	28.7

Abbreviations: CV, coefficient of variation; EIA, enzyme immunoassay; ELISA, enzyme-linked immunosorbent assay; IL, interleukin; multiplex bead, multiplex bead immunoassay; TNF-α, tumor necrosis factor-alpha; xMAP, magnetic bead–based multiplex immunoassay.

for IL-1, whereas variability between methods was much greater for TNF-α. We also noted that variability of cytokine measurements within a single laboratory, regardless of the method used tended to be moderately high with intralab CVs ranging from approximately 4% to close to 100% depending on the cytokine and the method.

When performing serial testing to monitor cytokine levels in a patient or subject, the same testing method should be used, ideally within the same laboratory. Using different methods for serial monitoring should not be used without careful comparison studies to validate this approach. Results reported by different laboratories whether using the same or a different method are not directly commutable and must be interpreted with caution. Even intralaboratory testing will routinely produce variability of up to 20%, and those interpreting changes in these results should recognize this limitation in precision when considering the clinical significance of serial cytokine measurements. Given the significant variability among cytokine assays and between laboratories, further efforts are required in order to standardize cytokine testing. This study was limited by the sample size and the potential role of using lyophilized samples in these analyses across multiple years. Additionally, the lyophilized material used in this study was initially analyzed by a single method by the supplier prior to distribution by the College of American Pathologists through the PT testing program. However, this is the largest comparison study published to date, and we hope this study serves to inform physicians, laboratorians, and investigators of the variability that is currently present in the laboratory measurement of cytokines.

The authors wish to thank Christine Bashleben for administrative assistance during development of this manuscript.

References

1. Dinarello CA. Historical insights into cytokines. *Eur J Immunol*. 2007;37 Suppl 1:S34–S45.
2. Krejci M, Janikova A, Folber F, Kral Z, Mayer J. Outcomes of 167 healthy sibling donors after peripheral blood stem cell mobilization with G-CSF 16mug/kg/day: efficacy and safety. *Neoplasma*. 2015;62(5):787–792.
3. Dale DC, Bonilla MA, Davis MW, et al. A randomized controlled phase III trial of recombinant human granulocyte colony-stimulating factor (filgrastim) for treatment of severe chronic neutropenia. *Blood*. 1993;81(10):2496–2502.
4. Jiang T, Zhou C, Ren S. Role of IL-2 in cancer immunotherapy. *Oncoimmunology*. 2016;5(6):e1163462.
5. Ma X, Xu S. TNF inhibitor therapy for rheumatoid arthritis. *Biomed Rep*. 2013;1(2):177–184.
6. Mukherjee M, Sehmi R, Nair P. Anti-IL5 therapy for asthma and beyond. *World Allergy Organ J*. 2014;7(1):32.
7. Nikolaus S, Waetzig GH, Butzin S, et al. Evaluation of interleukin-6 and its soluble receptor components sIL-6R and sgp130 as markers of inflammation in inflammatory bowel diseases. *Int J Colorectal Dis*. 2018;33(7):927–936.
8. Koga T, Migita K, Sato S, et al. Multiple serum cytokine profiling to identify combinational diagnostic biomarkers in attacks of familial mediterranean fever. *Medicine (Baltimore)*. 2016;95(16):e3449.
9. Lee DW, Gardner R, Porter DL, et al. Current concepts in the diagnosis and management of cytokine release syndrome. *Blood*. 2014;124(2):188–195.
10. Dranoff G. Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer*. 2004;4(1):11–22.
11. Tarrant JM. Blood cytokines as biomarkers of in vivo toxicity in preclinical safety assessment: considerations for their use. *Toxicol Sci*. 2010;117(1):4–16.
12. Aziz N, Nishanian P, Fahey JL. Levels of cytokines and immune activation markers in plasma in human immunodeficiency virus infection: quality control procedures. *Clin Diagn Lab Immunol*. 1998;5(6):755–761.
13. Aziz N, Nishanian P, Mitsuyasu R, Detels R, Fahey JL. Variables that affect assays for plasma cytokines and soluble activation markers. *Clin Diagn Lab Immunol*. 1999;6(1):89–95.