Conversion, Correction, and International Scale Standardization

Results From a Multicenter External Quality Assessment Study for *BCR-ABL1* Testing

Michael Griffiths, BSc; Simon J. Patton, PhD; Alberto Grossi, MD; Jordan Clark, MPhil; Maria Fe Paz, PhD; Emmanuel Labourier, PhD; on behalf of the Labceutics International BCR-ABL1 *Standardization Study Group*

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2013-0754-OA by guest on 28 October 2021

• *Context.*—Monitoring *BCR-ABL1* expression levels relative to clinically validated response criteria on the International Scale (IS) is vital in the optimal management of patients with chronic myeloid leukemia, yet significant variability remains across laboratories worldwide.

***Objective.*—To assess method performance, interlaboratory precision, and different IS standardization modalities in representative laboratories performing routine *BCR-ABL1* testing.**

***Design.*—Fifteen blinded test specimens with 5-level nominal *BCR-ABL1* to *ABL1* IS percentage ratios ranging from 5% to 0.0005% and 4-level secondary IS reference panels, the ARQ IS Calibrator Panels, were tested by relative quantitative polymerase chain reaction in 15 laboratories in 5 countries. Both raw and IS percentage ratios calculated by using local conversion factors (CFs) or analytic correction parameters (CPs) were collected and analyzed.**

***Results.*—A total of 670 valid positive results were generated. *BCR-ABL1* detection was associated with variable *ABL1* quality metric passing rates ($P < .001$) and reached at least 0.01% in 13 laboratories. Intralaboratory precision was within 2.5-fold for all sample levels combined with a relative mean difference greater than 5-fold across laboratories. International Scale accuracy was increased by using both the CF and CP standardization methods. Classification agreement for major molecular response status was 90% after CF conversion and 93% after CP correction, with precision improved by 3-fold for the CP method.**

***Conclusions.*—Despite preanalytic and analytic differences between laboratories, conversion and correction are effective IS standardization methods. Validated secondary reference materials can facilitate global diffusion of the IS without the need to perform sample exchange and improve the accuracy and precision of *BCR-ABL1* quantitative measurements, including at low levels of residual disease.**

(*Arch Pathol Lab Med.* 2015;139:522–529; doi: 10.5858/arpa.2013-0754-OA)

Molecular testing for the *BCR-ABL1* fusion gene by relative quantitative polymerase chain reaction (RQ-PCR) is currently the most sensitive routine approach for monitoring the response to therapy in patients with chronic myeloid leukemia (CML). In the context of tyrosine kinase inhibitor (TKI) therapies, the technique is most appropriate for patients who have achieved complete cytogenetic response and can be used to further define specific molecular response levels associated with therapeutic milestones.[1–4] Harmonization of the various laboratory procedures and reporting methods used worldwide is therefore essential to standardize optimal treatment response criteria and facilitate comparison across laboratories and patients.[5,6] After more than 10 years of continual efforts, it is now recognized that monitoring *BCR-ABL1* to control gene percentage ratios on the so-called International Scale (IS % ratios) is vital in the management of patients with CML and for optimizing outcomes.[7–9]

Attempts to harmonize procedures to measure the *BCR-ABL1* fusion transcripts e13a2 or e14a2 and monitor therapeutic response in patients with CML have included

Accepted for publication June 2, 2014.
Published as an Early Online Release July 25, 2014.
Supplemental digital content is available for this article at www.archivesofpathology.org in the April 2015 table of contents.

From West Midlands Regional Genetics Laboratory, Birmingham Women's NHS Foundation Trust, and School of Cancer Sciences, University of Birmingham, Birmingham, United Kingdom (Mr Griffiths); the European Molecular Genetics Quality Network, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester, United Kingdom (Dr Patton); Oncology Unit, Ospedale di Prato, Prato, Italy (Dr Grossi); United Kingdom National External Quality Assessment Schemes, Sheffield, United Kingdom (Mr Clark); Labceutics, Belfast, United Kingdom (Dr Fe Paz); and Asuragen, Austin, Texas (Dr Labourier). Mr Clark is now with Labceutics, Belfast, United Kingdom.

This work was supported by Asuragen (Austin, Texas). Labceutics (Belfast, United Kingdom) was commissioned by Asuragen to undertake the study in order to ensure the study was carried out independently. Mr Griffiths, Dr Patton, and Dr Grossi received honoraria from Labceutics for acting as advisors to the study. Dr Fe Paz is employed by Labceutics and Dr Labourier is employed by Asuragen. Mr Clark had no relevant financial interest in the products or companies described in this article at the time this study was prepared.

Reprints: Emmanuel Labourier, PhD, Asuragen, 2150 Woodward St, Suite 100, Austin, TX 78744 (e-mail: elabourier@asuragen.com).

significant investments in sample exchange. Converting *BCR-ABL1* expression levels measured in local clinical laboratories to the IS via laboratory-specific conversion factors (CFs) has been shown to improve major molecular response (MMR or $MR^{3.0}$) concordance rates in approximately 50% of laboratories.[10,11] This approach is however limited by the recognized interlaboratory variation in *BCR-ABL1* quantitative measurements and the lack of a common set of reference samples that can be shared on a global scale. This need was recently addressed by the formulation and validation of the first World Health Organization (WHO) International Genetic Reference Panel for quantitation of *BCR-ABL1* messenger RNA by RQ-PCR.[12,13] The WHO primary standards consist of a 4-level panel of e14a2-positive lyophilized cell line dilutions. Each level has an assigned IS reference value, the mean IS % ratio that was obtained by repeated testing of each sample level in expert IS-standardized laboratories during the establishment of these materials.[13]

Approximately 3500 primary reference panels were initially manufactured. To ensure the long-term supply of stable, homogeneous, commutable, and traceable reference standards to hundreds of clinical laboratories worldwide, secondary reference panels were developed and validated.[14] The panels consist of synthetic, nuclease-resistant Armored RNA Quant (ARQ) e13a2 or e14a2 fusion transcripts serially diluted in a background of ARQ *ABL1* and *BCR* transcripts with lot-specific IS % ratios anchored to the consensus reference values of the WHO primary standards. A broad validation study demonstrated that the 4-level ARQ IS Calibrator Panels are reproducibly manufactured and compatible with the accurate alignment of local RQ-PCR methods to the IS through derivation of laboratory-specific analytic correction parameters (CPs).[14] Because of their design enabling assessment of assay linearity, sensitivity, and precision, it was also proposed that the secondary IS reference materials may be used in interlaboratory comparative studies or external quality assessment (EQA) programs.

The aim of the present study was to conduct an international, multicenter, EQA-modeled scheme by using both 5-level, blinded test samples and the ARQ IS Calibrator Panels. The study provided a platform on which to assess the performance of 15 distinct clinical *BCR-ABL1* tests, compare different IS standardization modalities, and confirm the utility of secondary reference materials to further improve IS accuracy and interlaboratory precision.

MATERIALS AND METHODS

Study Design and Oversight

The study was designed and conducted by Labceutics (Belfast, United Kingdom) with input from an advisory panel of experts. Clinical laboratories routinely performing *BCR-ABL1* testing (n = 15) were enrolled via open call in 5 countries in Europe. Blinded test samples were provided by the United Kingdom National External Quality Assessment Schemes for Leucocyte Immunophenotyping (UK NEQAS LI, Sheffield, United Kingdom). The WHO primary standards (NIBSC code 09/138) were obtained from the United Kingdom National Institute for Biological Standards and Control (Potters Bar, United Kingdom). The synthetic ARQ IS Calibrator Panels with lot-specific IS % ratios anchored to the WHO primary standards (Supplemental Table 1, see supplemental material file at www.archivesofpathology.org in the April 2015 table of contents) were provided by Asuragen (Austin, Texas). Raw data generated by the 15 participating laboratories were directly downloaded by each site to Labceutics' secured Web-based server. The primary data analysis was performed in a double-blinded way with review and input from the advisory panel. The international study group met to review and discuss the data during a half-day consensus workshop held on June 12, 2013, in Stockholm, Sweden, where participants agreed to submit the data for publication. All authors reviewed the manuscript and approved the decision to submit for publication.

Cell Line Dilution Samples

Five batches of test samples were prepared and qualified by UK NEQAS LI according to standard operating procedures. The samples were produced by serial dilution of the t(9;22)-positive cell line K562 (expressing *BCR-ABL1* e14a2 fusion transcript) in the t(9;22)-negative cell line HL-60. Formulations were designed to target IS % ratios below each of the established TKI clinical response criteria, that is, between 10% (complete hematologic response), 1% (complete cytogenetic response), 0.1% (MMR or $MR^{3.0}$), 0.01% ($MR^{4.0}$), and 0.001% ($MR^{5.0}$), and with the last dilution expected to be negative for *BCR-ABL1* or beyond the quantitative range of most RQ-PCR methods. The initial dilution was designed to produce an IS % ratio of approximately 5%. Samples were stabilized by lyophilization, stored below −15°C, and shipped at room temperature with temperature-monitoring devices.[15] Each batch of samples was labeled in a blinded fashion such that participating laboratories received 3 vials from each of the 5 batches of samples (15 blinded samples total per laboratory).

Molecular Methods

Quality control on 3 cell line samples randomly selected from each of the 5 batches was performed by spectrophotometry, electrophoresis (Bioanalyzer 2100, Agilent Technologies, Santa Clara, California), and RQ-PCR per UK NEQAS LI's standard operating procedures. Nominal IS % ratios were assigned to each cell line level by testing one set of 15 blinded samples in an independent laboratory (Asuragen Clinical Laboratory, Austin, Texas) with the BCR/ABL1 Quant Kit[16] (Asuragen) and comparing the measured percentage ratios to the reference IS % ratios of the primary or secondary standards as previously described.[13,14] At each participating laboratory, the cell line dilution samples were extracted by using laboratory-specific protocols and following recommendations for the handling of lyophilized cell lines provided by UK NEQAS LI. The ARQ IS Calibrator Panels were tested directly after heat lysis at 75°C for 5 minutes. All RQ-PCR tests were performed according to procedures specific to each of the 15 participating laboratories, including for number of replicates and quality control criteria. Most laboratories used a minimum mean copy number requirement of 10 000 copies of *ABL1* per RQ-PCR, 4 used other *ABL1* cutoffs (32 000 copies, n = 1; 5000 copies, n = 2; 1000 copies, n = 1), and 1 used a cycle threshold (Ct) cutoff (*ABL1* Ct < 30). The molecular methods were based on commercial reagents (n = 5) or laboratory-developed protocols (n = 10, including 5 based on the Europe Against Cancer protocol[17,18]) and were performed on a variety of real-time PCR instruments (Life Technologies, Carlsbad, California, n = 9; Roche, Basel, Switzerland, n = 5; Agilent Technologies, n = 1).

Data Analysis

Laboratory-specific CPs were calculated as the antilog of the arithmetic mean of the differences between log-transformed reference IS % ratios (Supplemental Table 1) and log-transformed raw percentage ratios measured in each laboratory for each ARQ IS Calibrator Panel and level, that is, $CP = 10^{\frac{1}{n}\sum_{i=1}^{n}(\log_{10} IS_i - \log_{10} Raw_i)}$.[14] *BCR-ABL1* detection rates and quality control passing rates for both the Calibrators and the blinded cell line samples were calculated for each laboratory by using all RQ-PCR results submitted. Bias and correlation analyses were performed by using the Bland-Altman method and least squares linear regression analyses as previously described.[10,13,14] Variability at individual sample levels, across multiple sample levels, and in pooled analyses was assessed by

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2013-0754-OA by guest on 28 October 2021

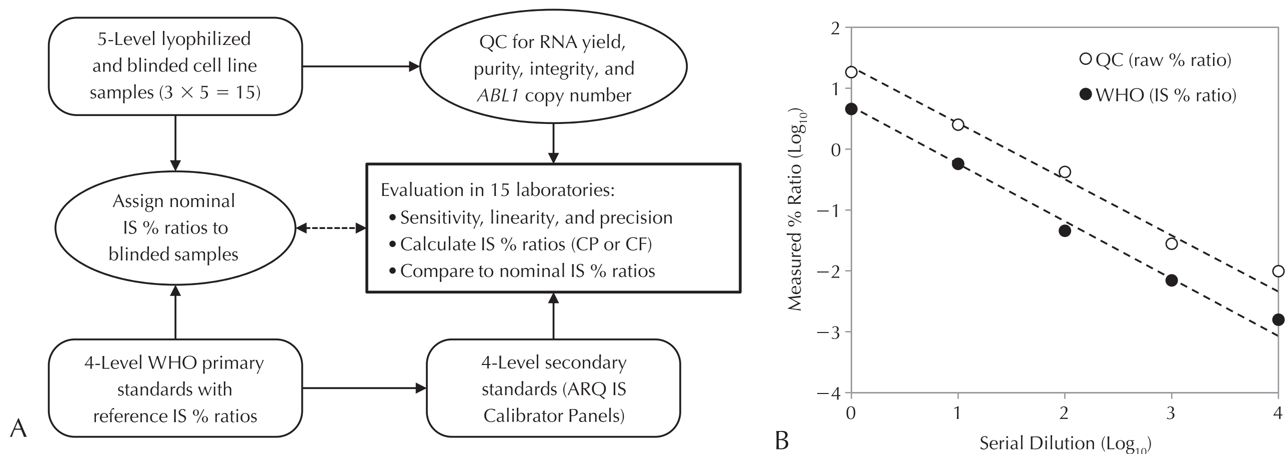

Figure 1. *Study overview. A, Study design. Blinded cell line dilutions (e14a2) and ARQ IS Calibrator Panels (e13a2 or e14a2) were tested in 15 laboratories.* BCR-ABL1 *to* ABL1 *percentage ratios generated in each laboratory were used to assess method performance and IS standardization. B,* BCR-ABL1 *to* ABL1 *percentage ratios measured during quality control testing at the United Kingdom National External Quality Assessment Schemes for Leucocyte Immunophenotyping (QC, raw % ratio) or by comparison to the WHO primary standards (WHO, IS % ratio). The graph shows the mean percentage ratios measured for sample levels 1 to 5 (10-fold serial dilutions) after log base 10 transformation. The dashed lines represent the least squares linear fit between the 4 first sample levels. Abbreviations: ARQ, Armored RNA Quant; CF, conversion factor; CP, correction parameter; IS, International Scale; QC, quality control; WHO, World Health Organization.*

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2013-0754-OA by guest on 28 October 2021

using log base 10–transformed percentage ratios before calculation of the 95% limits of agreement (95% LOA). Percentage ratio values are not normally distributed and coefficients of variation (CV) for percentage ratios at individual sample levels are reported in a supplemental file for information purposes only. The MMR classification accuracy was estimated by using standard statistics for binary qualitative molecular test results (positive = in MMR = below 0.1% IS, negative = above 0.1% IS) and the exact Clopper-Pearson method for 95% confidence intervals. To avoid interference by inaccurate or imprecise measurements that can bias and/or mask statistically significant results, all pooled analyses were performed after exclusion of percentage ratios from laboratories with abnormally low *ABL1* passing rates (<50%) and *BCR-ABL1* detection rates (<95% for sample levels 1 to 3) or abnormally high variability (95% LOA > 5-fold). When indicated, *P* values were calculated by using the Fisher exact test for categorical variables or linear regression for continuous variables.

RESULTS

Study Design and Samples

Blinded, lyophilized cell line samples (e14a2, 5 dilution levels) and ARQ IS Calibrator Panels (e13a2 or e14a2, 4 levels) were tested in 15 clinical laboratories in France, Germany, Italy, Spain, and the United Kingdom (Figure 1, A). Cell line dilutions were formulated to target IS % ratios from below 10% to below 0.001%, with the last dilution expected to be negative for *BCR-ABL1* or beyond the quantitative range of most RQ-PCR methods. Raw *BCR-ABL1* to *ABL1* percentage ratios generated in each laboratory were used to assess the sensitivity, linearity, and precision of local RQ-PCR methods. Following conversion of raw percentage ratios to IS % ratios using laboratory-specific CF and/or CP, interlaboratory variability and IS standardization were assessed relative to nominal IS % ratios independently assigned to each cell line level by direct comparison to the WHO primary standards.

Before dispatching, the lyophilized cell line samples met all predefined quality control criteria for RNA yield, purity, and integrity and minimum *ABL1* copy number per PCR (Table 1). Raw percentage ratios were consistent with a 10-fold serial dilution and were linear for the first 4 levels (Figure 1, B). As expected, *BCR-ABL1* was inconsistently detected below the linear quantitative detection range of the RQ-PCR method in the lowest positive samples (level 5; samples E, F, and I). Direct comparison to the reference IS % ratios of the WHO primary standards resulted in nominal IS % ratios ranging from approximately 5% to 0.001% with a similar loss of linearity for samples E, F, and I (Figure 1, B; Table 1). Based on the 10-fold dilution between each sample level, rounded nominal IS % ratios of 5%, 0.5%, 0.05%, 0.005%, and 0.0005% were assigned to

Table 1. Summary of Quality Control Results and International Scale (IS) Percentage Ratio Assignment Using the World Health Organization Primary Standards (WHO) or the Secondary ARQ IS Calibrator Panels (ARQ)

		Quality Control at UK NEQAS LI							Assignment of IS% Ratios		
Level	Sample	Yield, g	260/280 OD Ratio	260/230 OD Ratio	RIN	rRNA Ratio	ABL1 (cp/PCR)	Raw % Ratio	WHO	ARQ	Nominal
1	A H N	36.3	2.00	2.12	7.7	2.0	24 337	18.4	4.6	6.7	5
2	B L M	33.5	2.01	2.13	7.4	2.6	20 104	2.5	0.57	0.74	0.5
3	C D K	34.7	2.01	2.15	7.7	2.1	25 537	0.42	0.046	0.051	0.05
4	G J O	34.8	2.00	1.99	7.1	2.1	37 444	0.028	0.0070	0.0070	0.005
5	E F I	34.5	2.00	2.12	7.3	2.1	32 752	0.0099[a]	0.00158[a]	0.00146[a]	0.0005

Abbreviations: ARQ, Armored RNA Quant; cp/PCR, copies per polymerase chain reaction; OD, optical density; RIN, RNA integrity number; rRNA, ribosomal RNA; UK NEQAS LI, United Kingdom National External Quality Assessment Schemes for Leucocyte Immunophenotyping.

[a] Positive but below linear quantitative detection range of assay.

Table 2. Summary of Evaluation Results With Blinded Cell Line Samples										
		BCR-ABL1 Detection Rate in Sample Levels, %							Relative Bias	
	Pass *ABL1*, %	1	2	3	4	5	1–5	1–4	Fold	95% LOA
Lab 3	3.3	100	100	83	0	0	57	71	. . .	. . .
Lab 4	100	100	100	100	89	44[a]	87	97	−2.59	1.96
Lab 5	80	100	100	100	100[a]	0	80	100	1.02	1.48
Lab 10	100	100	100	100	89[a]	22[a]	82	97	−1.16	1.55
Lab 15	93	100	100	100	100	33[a]	87	100	−2.61	1.96
Lab 18	0	100	100	83	50[a]	0	67	83	. . .	. . .
Lab 20	91	100	100	100	100	0	80	100	−5.94	1.79
Lab 22	100	100	100	100	100	33[a]	87	100	−2.38	2.34
Lab 24	53	100	100	100	100[a]	0	80	100	1.26	1.81
Lab 25	0	100	78	11	0	0	38	47	. . .	. . .
Lab 32	97	100	100	100	100	17[a]	84	100	1.41	5.97
Lab 40	0	100	67	83	17[a]	0	53	67	. . .	. . .
Lab 41	84	100	100	100	100[a]	0	81	100	1.14	2.08
Lab 42	0	100	0	0	0	0	20	25	. . .	. . .
Lab 100	10	100	100	50	0	0	50	58	. . .	. . .

Abbreviations: Lab, laboratory; 95% LOA, 95% limits of agreement.

[a] Positive but below linear quantitative detection range of assay.

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2013-0754-OA by guest on 28 October 2021

sample levels 1 to 5 and used for all subsequent analyses. As a control, the blinded samples were also compared to the reference IS % ratios of the ARQ IS Calibrator Panels. The nominal IS % ratios obtained with the primary and secondary IS standards were in the same range (Table 1) with a mean bias of 1.15-fold and 95% LOA of ±1.41-fold (Supplemental Figure 1).

Evaluation of Blinded Samples

A total of 318 *BCR-ABL1* positive results were generated in cell line dilution samples that met laboratory-specific quality control criteria (Supplemental Figure 2). Three laboratories obtained *ABL1* copy numbers above their respective minimum copy number requirements for all samples evaluated in all runs performed (Table 2). In 6 laboratories, less than 10% of the PCR amplifications resulted in reportable percentage ratios with sufficient *ABL1* copy numbers. In these laboratories, the overall *BCR-ABL1* detection rate was 20% to 67% and none could reproducibly detect *BCR-ABL1* in samples that mimicked a clinical specimen at or just below MMR (level 3; samples C, D, and K; 0.05% on the IS). For the 9 other laboratories, the overall *BCR-ABL1* detection rate ranged from 80% to 87% and was 97% to 100% after exclusion of the lowest positive samples E, F, and I, which were expected to return negative results (sample levels 1–4; Table 2). Five laboratories could however detect *BCR-ABL1* in 17% to 44% of the amplification reactions with level-5 samples, each time outside the linear quantitative detection range of their respective RQ-PCR methods. *BCR-ABL1* was reproducibly detected in level-4 samples (samples G, J, and O) but was outside the local method's linear range in 4 laboratories (Table 2).

For the 9 laboratories that obtained positive results over at least 3 logs of cell line dilution, all raw percentage ratios measured within the linear range of each local RQ-PCR method were compared against a single set of reference values, the nominal IS % ratios independently assigned to each sample level (5% to 0.005%). The relative mean bias ranged from –5.94-fold to +1.41-fold in distinct laboratories (Table 2), indicating a broad interlaboratory variability. The intralaboratory precision was good with an average 95% LOA of 2.3-fold (Table 2). The 95% LOA were less than 2.5-fold in most cases with corresponding level-specific CV between 7% and 73% (mean CV of 27%, 34%, 44%, or 52% for sample levels 1 to 4; Supplemental Table 2). The single laboratory with 95% LOA greater than 2.5-fold had higher CVs and a low coefficient of determination for the corresponding least squares linear regression analysis ($R^2 = 0.884$; Supplemental Figure 2).

Evaluation of ARQ IS Calibrator Panels

The e13a2 and e14a2 secondary reference panels passed laboratory-specific quality control criteria in 359 of 360 tests performed (99.7%) and *BCR-ABL1* was detected in 353 amplification reactions (98%), providing 352 valid data points (Supplemental Figure 3, A). The lowest e13a2 positive Calibrator (level 4) was not detected in 1 run by a single laboratory, while the e14a2 Calibrator level 4 was not detected in any run by 2 distinct laboratories (Table 3). For the 13 laboratories that detected all Calibrator levels, laboratory-specific analytic CPs were derived by comparing the local raw percentage ratios to the lot-specific IS % ratios assigned to each Calibrator level (from approximately 4% to below 0.01%; Supplemental Table 1). For the 2 laboratories that did not reproducibly detect levels 4, CP values were estimated by using only 3 Calibrator levels and are also reported in Table 3 for completeness. The CPs varied widely from 0.25-fold to 5.18-fold, again confirming a broad interlaboratory variability (Table 2). The relative difference between CPs derived from the e13a2 or e14a2 panels was negligible in most laboratories (≤1.5 in 11 laboratories, ≤2.0 in 13 laboratories), suggesting similar assay performance characteristics for both e13a2 and e14a2 fusion transcripts. The highest relative CP difference (3.4) was observed for laboratory 24 where e13a2, but not e14a2, was detected in Calibrators level 4.

For most laboratories, the intralaboratory variability was within the expected range previously established in precision and multisite evaluation studies using the ARQ IS Calibrator Panels.[14] The 95% LOA were 2.0-fold on average and were 2.50-fold or less in 12 laboratories (Table 3) with level-specific CVs between 10% and 83% (Supplemental Table 2). The 95% LOA were greater than 2.50-fold in a single laboratory for the e13a2 panel and in 3 laboratories for the e14a2 panel (Table 3). Lower precision for both the e13a2 and e14a2 fusion transcripts was also associated with

Table 3. Summary of Evaluation Results With ARQ IS Calibrator Panels								
		e13a2 Panel			e14a2 Panel			
	Pass *ABL1*, %	Pos, %	CP	95% LOA	Pos, %	CP	95% LOA	Relative CP Difference
Lab 3	100	100	3.59	1.35	100	3.65	3.57	1.0
Lab 4	100	100	0.51	1.61	100	0.49	1.85	1.1
Lab 5	100	100	2.08	1.91	100	0.84	2.50	2.5
Lab 10	100	100	1.14	1.66	100	0.97	2.02	1.2
Lab 15	100	100	1.14	1.34	100	1.13	1.75	1.0
Lab 18	100	100	1.41	1.90	100	1.95	1.99	1.4
Lab 20	100	100	0.30	2.25	100	0.25	3.08	1.2
Lab 22	100	100	0.51	1.59	100	0.67	1.78	1.3
Lab 24[a]	100	100	1.53	2.05	75	5.18	2.01	3.4
Lab 25[b]	96	92	0.38	1.40	75	0.50	1.58	1.3
Lab 32	100	100	4.08	1.84	100	2.75	1.99	1.5
Lab 40	100	100	0.72	3.41	100	0.63	2.50	1.1
Lab 41	100	100	1.61	2.12	100	0.79	1.95	2.0
Lab 42	100	100	0.55	2.11	100	0.36	2.54	1.5
Lab 100	100	100	3.67	1.94	100	2.22	1.84	1.7

Abbreviations: ARQ, Armored RNA Quant; CP, correction parameter; IS, International Scale; Lab, laboratory; Pos, positive for *BCR-ABL1*; 95% LOA, 95% limits of agreement.

[a] *BCR-ABL1* not detected in e14a2 Calibrator 4 (3 of 3 runs).
[b] *BCR-ABL1* not detected in e13a2 (1 of 3 runs) and e14a2 (3 of 3 runs) Calibrator 4.

a suboptimal linear fit in least squares regression analyses (Supplemental Figure 3, B and C). Following correction of the raw percentage ratios with laboratory-specific CP, the bias was null and corrected IS % ratios were within ±2-fold, as expected (Supplemental Figure 3, D).

IS Standardization Across Laboratories

A total of 8 laboratories reported valid *BCR-ABL1* quantitative measurements over at least 3 logs of cell line dilution and had an established laboratory-specific CF. Bias analyses between the nominal IS % ratios anchored to the WHO primary standards and the local measured IS % ratios after CF conversion (IS_{CF} % Ratio = CF × Raw % Ratio) resulted in a mean bias ranging from approximately –5-fold to +5-fold in distinct laboratories (Figure 2, A). The intralaboratory variability is independent of the IS standardization process and remained ±1.48-fold for laboratory 5 to ±5.97-fold for laboratory 32 (Figure 2, A; Table 2). There were also 8 laboratories with valid *BCR-ABL1* quantitative measurements in lyophilized cell line samples and valid analytic CPs. Following correction of local raw percentage ratios with laboratory-specific e14a2 CP (IS_{CP} % Ratio = CP × Raw % Ratio), the distribution of mean biases was tighter with values ranging from approximately –3-fold to +1.5-fold (Figure 2, A). The overall 95% LOA for all 8 laboratories combined, calculated by using mean IS % ratios for each sample level in each laboratory, that is, after removal of the inherent intralaboratory variability, was almost 3 times higher for CF conversion (±7.65-fold) than for CP correction (±2.81-fold) ("All", Figure 2, A).

A total of 7 laboratories had both a valid CF and a valid CP, enabling direct comparison of the 2 IS standardization methods. Bias analyses using all valid percentage ratios generated in these laboratories (n = 151) confirmed that both methods improved the accuracy of *BCR-ABL1* quantitative measurements with a variability approximately 2.5-fold higher for the CF conversion method (Supplemental Figure 4, A). After exclusion of the laboratory with an uncharacteristic high imprecision (laboratory 32, 95% LOA of 5.97; Table 2), the variability of CF-converted IS % ratios was ±8.1- to 8.7-fold at different cell line dilution levels versus ±2.3- to ±3.1-fold for CP-corrected IS % ratios (Figure 2, B). For all levels combined, the accuracy in *BCR-ABL1* quantitative measurements was improved by approximately 2.1-fold on average after CF conversion with a variability of ±9.2-fold and was improved by approximately 1.6-fold after CP correction with a variability of ±2.9-fold (Figure 2; Supplemental Figure 4). The accuracy of MMR classification was high for both methods with overlapping 95% confidence intervals suggesting similar performance (Table 4). With the CP method, 57 of 151 valid test results were classified in MMR (IS % ratio < 0.1%) and 93% (141 of 151) of the results were in qualitative agreement with the nominal IS % ratios anchored to the WHO primary standards. With the CF method, 55 test results were classified in MMR with 90% agreement (136 of 151) relative to the nominal values.

COMMENT

Despite the proven clinical utility of *BCR-ABL1* monitoring and significant efforts over many years to improve the standardization of *BCR-ABL1* quantitative measurements, substantial variability remains across clinical laboratories worldwide.[8,9] The establishment and validation of consensus IS reference values through the WHO primary standards and secondary reagents was undoubtedly a significant milestone for IS standardization.[13,14] In the present study, we further show that synthetic IS reference materials anchored to the WHO primary standards can reduce interlaboratory variability. The accuracy of *BCR-ABL1* measurements after CP correction, using a single lot of reagent, was similar to that obtained with the current CF method, with a 3-fold increase in precision. Our results substantiate the objectives initially set during the establishment of the primary WHO reference standards, that is, to facilitate worldwide diffusion of the IS and improve the quality of molecular data that are critical for optimal patient management.[7,13]

Voluntary participation in proficiency surveys and other EQA programs is an important tool toward achieving comparability of measurement between laboratories. It is also valuable for clinical laboratories to identify potential performance issues and engage in corrective or preventive improvement activities. In our study, there was a statistically

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2013-0754-OA by guest on 28 October 2021

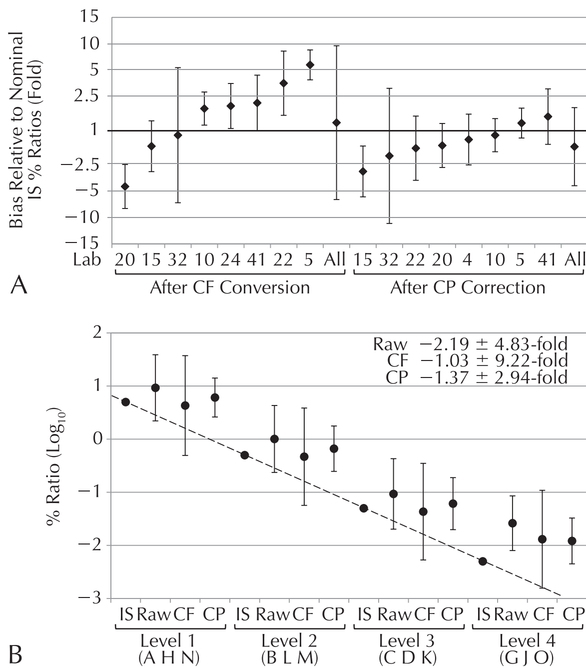

Figure 2. *Accuracy and precision of IS standardization. A, Relative difference between nominal and measured IS % ratios. The mean bias (black diamond) was calculated by using CF-converted or CP-corrected raw percentage ratios for all cell line samples detected over at least 3 levels of dilution within the linear range of each local RQ-PCR method in the indicated laboratories. The overall bias for all laboratories combined ("All") was calculated after averaging the IS % ratios for each cell line level in each laboratory. The error bars represent the 95% LOA. B, Average nominal (IS), raw (Raw), CF-converted (CF), and CP-corrected (CP) percentage ratios per cell line level. The graph shows the average (black circle) and standard deviation (error bar) for all percentage ratios obtained in laboratories 5, 10, 15, 20, 22, and 41 at each sample level. The corresponding mean bias and 95% LOA calculated by using all measured percentage ratios at all dilution levels combined are shown in the top right corner. The dashed line represents the theoretic perfect linear fit between the nominal IS % ratios assigned to each of the 4 sample levels (5% to 0.005%). Abbreviations: CF, conversion factor; CP, correction parameter; IS, International Scale; Lab, laboratory; RQ-PCR, relative quantitative polymerase chain reaction; 95% LOA, 95% limits of agreement.*

significant association ($P < .001$) between the percentage of tests with *ABL1* copy numbers passing laboratory-specific quality criteria and the *BCR-ABL1* detection rates in sample level 3 (samples C, D, and K; 0.05% on the IS) or in sample levels 1 to 4 combined (5% to 0.005% on the IS). Overall, 9 of 15 laboratories (60%) could reproducibly detect *BCR-ABL1* below MMR in every test performed with samples C, D, and K, and 7 laboratories (47%) could reproducibly detect *BCR-ABL1* below $MR^{4.0}$ in every test performed with samples G, J, and O. These low detection rates reflect unexpected difficulties and variations during RNA extraction from lyophilized cell line pellets and are likely not representative of the analytic or clinical sensitivity of the RQ-PCR methods used by the participating laboratories. This was confirmed during evaluation of the secondary IS reference standards that are designed to prevent RNA loss or degradation during molecular testing and to challenge the analytic sensitivity of local RQ-PCR methods.[14] With these materials, 13 laboratories (87%) reproducibly detected *BCR-ABL1* below $MR^{4.0}$ with a reported mean copy number of 12 copies per RQ-PCR test.

Table 4. Evaluation of Major Molecular Response Classification Accuracy

Reference IS % Ratios	Comparator IS % Ratios	Agreement (95% CI), %
Nominal (WHO)	CP corrected	93 (88–97)
Nominal (WHO)	CF converted	90 (84–94)
CF converted	CP corrected	87 (81–92)

Abbreviations: CF, conversion factor; CP, correction parameter; IS, International Scale; WHO, World Health Organization; 95% CI, 95% confidence interval.

There was also a noticeable association between laboratories with an established CF and higher *BCR-ABL1* detection rates in the blinded cell line samples. In 8 of 9 laboratories that reported a CF, enough valid positive results were generated across at least 3 logs of cell line dilution to perform interlaboratory and intralaboratory precision analyses, while this was achieved by only 1 of 6 laboratories not IS standardized ($P = .01$). This observation suggests that experience with standardization procedures, quality assessment programs, and/or various sample types in IS-standardized laboratories may result in improved performance during the preanalytic and analytic steps and in achieving higher apparent analytic sensitivity. A similar trend was observed in laboratories participating in the 2012 *BCR-ABL1* proficiency survey organized by the College of American Pathologists where 74% of IS laboratories and 62% of non-IS laboratories could detect *BCR-ABL1* below $MR^{4.0}$.[19] One limitation in our study, however, is that the number of participating laboratories was limited to 15 in 5 European countries. Additional studies in other countries, preferably sponsored by independent professional medical bodies and using multiple lots of Calibrators, could further our understanding of the benefits of IS standardization by either the CF or CP methods. This may be particularly relevant in North America where *BCR-ABL1* testing is decentralized and where only 24% of clinical laboratories participating in proficiency surveys are standardized through sample exchange with IS reference laboratories.[19]

Analytic differences between independent measurement methods are inevitable. Even after IS standardization, the process cannot be mathematically perfect considering the breadth of RNA extraction and cDNA (complementary DNA) generation protocols, quantitative technologies and platforms, assay designs, standard curves materials, and reporting methods used by clinical laboratories worldwide. Previous studies have shown that both bias and imprecision contribute to interlaboratory variability, with a mean bias of ±1.2-fold and 95% LOA of ±5-fold expected across IS laboratories.[10,11] In our study, the mean bias was reduced to less than 1.2-fold after CF conversion but with a variability of ±9-fold for IS % ratios measured in distinct laboratories. This result is not surprising since the above-referenced studies used a single reference method in a single IS laboratory, while in the clinical setting and in our study, CFs are established by sample exchange with different IS reference laboratories in different countries. In addition, the nominal IS % ratios assigned to our blinded cell line samples were directly anchored to the reference values of the WHO primary standards. Those reference values represent the average *BCR-ABL1* to *ABL1* IS % ratios measured in 6 experienced IS laboratories with a reported variability of up to 4-fold between the 6 laboratories.[13] When the laboratories participating in our study were

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2013-0754-OA by guest on 28 October 2021

standardized against a unique set of IS reference values through the ARQ IS Calibrator Panels, both the bias and variability were decreased with 95% LOA of ±3-fold, as expected from previous multisite validation studies.[14] The level-specific interlaboratory CVs in the first 4 levels of lyophilized cell line samples were 37% to 54% for CP-corrected IS % ratios, which is similar to the performance reported during the establishment of the WHO primary standards (interlaboratory CVs of 32% to 41% for the 4-level lyophilized cell line panels).[13]

Regardless of the analytic variability in quantitative measurements of *BCR-ABL1*, it is important to keep in mind that the purpose of laboratory improvement programs is to achieve medical usefulness. The ultimate goal of IS standardization is not to obtain the exact same analytic values in all laboratories but to harmonize the definition of optimal treatment response criteria that have been validated in rigorous drug clinical trials.[3,4,6,20] Clinically, molecular response to TKIs is assessed relative to MMR, a 3-log reduction from the mean percentage ratio that was obtained by testing a common set of clinical specimens at diagnosis during the International Randomized Study of Interferon and STI571 trial and defined as 100% IS.[6,20,21] Previous studies have shown that because of the inherent limitations and variability of current RQ-PCR technologies (intralaboratory precision of ±2.5-fold), an optimal MMR classification agreement of 88% to 95% is expected across IS-standardized laboratories.[10,11] An overall classification accuracy of 90% was previously reported for the ARQ IS Calibrator Panels with 92% agreement between the CF and CP methods.[14] In our study, the intralaboratory precision was 2.3-fold on average, the MMR classification accuracy for CP-corrected IS % ratios was 93%, and the classification agreement between CP-corrected and CF-converted IS % ratios was 87% across a clinically representative range of measurements. The misclassified samples generated IS % ratios within 1.5-fold of MMR on average with 95% LOA of ±1.9-fold (range of IS % ratios, 0.08%–0.4%), that is, they were all within the expected intralaboratory variability for the quantitative measurement of *BCR-ABL1* to *ABL1* percentage ratios around MMR (0.1% on the IS).

At lower *BCR-ABL1* expression levels, the classification accuracy was similar with 92% agreement between CP-corrected and CF-converted IS % ratios in samples classified above or below 0.01% (95% confidence intervals, 87%–96%). As prolonged molecular responses greater than 4 logs on the IS ($MR^{4.0}$) are more commonly observed with second-generation TKIs, the definition and interpretation of these extremely low to undetectable levels of residual disease will need to be further standardized in the near future.[21,22] Our study has shown that CPs derived from the ARQ IS Calibrator Panels can provide an effective approach to IS standardization, allowing laboratories to deliver results on the IS below MMR and $MR^{4.0}$ without the need for relatively lengthy and expensive sample exchanges associated with the derivation of CFs. We conclude that inclusion of secondary IS reference panels in routine clinical testing, external quality assessment programs, and proficiency surveys can not only further the analytic calibration of local RQ-PCR methods to the IS but also improve the accuracy of clinical measurements in patients with CML who have achieved a deep and durable molecular response. The availability of these validated secondary reference materials currently benefits clinical laboratories, diagnostic test developers, pharmaceutical companies, and medical professionals by providing more reliable molecular data that ultimately will further improve the management of patients with CML.

The Labceutics International *BCR-ABL1* Standardization Study Group consists of Susanna Akiki, PhD (WMRGL, Birmingham Women's NHS Foundation Trust, Birmingham, United Kingdom), Rosa Ayala, MD, PhD (Hospital Doce de Octubre, Madrid, Spain), Eva Barragan, PhD (Hospital La Fé, Valencia, Spain), Daniela Basso, MD (Department of Laboratory Medicine, University Hospital of Padova, Italy), Nathalie Beaufils, MD (CHU-CLCC de Marseille, Marseille, France), Dominique Bories, MD, PhD (CHU Henri Mondor, Paris, France), Jean-Michel Cayuela, PharmD, PhD (Laboratory of Hematology, University Hospital Saint-Louis, Paris, France), Jordan Clark, MPhil (UK NEQAS, Sheffield, United Kingdom), Marc Füllgrabe, PhD (Universitätsklinikum Schleswig-Holstein, Kiel, Germany), Jean Gabert, MD, PhD (CHU-CLCC de Marseille, Marseille, France), Michael Griffiths, BSc (WMRGL, Birmingham Women's NHS Foundation Trust, Birmingham, United Kingdom), Alberto Grossi, MD (Ospedale di Prato, Prato, Italy), Aytug Kizilors, PhD (King's College Hospital, London, United Kingdom), Emmanuel Labourier, PhD (Asuragen, Austin, TX, USA), Nicholas Lea, PhD (King's College Hospital, London, United Kingdom), Luis Lombardia, PhD (CNIO, Madrid, Spain), Carole Maute, MSc (Laboratory of Hematology, University Hospital Saint-Louis, Paris, France), Sarah L. McCarron, MSc (St. James Hospital, Dublin, Ireland), Guillermina Nickless, MSc (Guys and Saint Thomas Hospital, London, United Kingdom), Simon J. Patton, PhD (EMQN, Manchester, United Kingdom), Maria Fe Paz, PhD (Labceutics, Belfast, United Kingdom), Elisa Piva, MD (Department of Laboratory Medicine, University Hospital of Padova, Italy), Christiane Pott, MD (Universitätsklinikum Schleswig-Holstein, Kiel, Germany), Cristina Rabascio, BSc (European Institute of Oncology, Milan, Italy), Daniel Rueda, PhD (Hospital Doce de Octubre, Madrid, Spain), Christoph Schmitt, PhD (IMOGEN GmbH, Bonn, Germany), Oliver Wachter, BTA (Zentrum fuer Humangenetik und Laboratoriumsmedizin, Martinsried, Germany), and Lurdes Zamora, PhD (ICO-Hospital Germans Trias i Pujol, Instituto de Investigación contra la Leucémia Josep Carreras, Badalona, Spain).

References

1. Hanfstein B, Muller MC, Hehlmann R, et al. Early molecular and cytogenetic response is predictive for long-term progression-free and overall survival in chronic myeloid leukemia (CML). *Leukemia*. 2012;26(9):2096–2102.

2. Marin D, Ibrahim AR, Lucas C, et al. Assessment of BCR-ABL1 transcript levels at 3 months is the only requirement for predicting outcome for patients with chronic myeloid leukemia treated with tyrosine kinase inhibitors. *J Clin Oncol*. 2012;30(3):232–238.

3. Hochhaus A, O'Brien SG, Guilhot F, et al. Six-year follow-up of patients receiving imatinib for the first-line treatment of chronic myeloid leukemia. *Leukemia*. 2009;23(6):1054–1061.

4. Hughes TP, Kaeda J, Branford S, et al. Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. *N Engl J Med*. 2003;349(15):1423–1432.

5. Branford S, Cross NC, Hochhaus A, et al. Rationale for the recommendations for harmonizing current methodology for detecting BCR-ABL transcripts in patients with chronic myeloid leukaemia. *Leukemia*. 2006;20(11):1925–1930.

6. Hughes T, Deininger M, Hochhaus A, et al. Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results. *Blood*. 2006;108(1):28–37.

7. Baccarani M, Deininger MW, Rosti G, et al. European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. *Blood*. 2013;122(6):872–884.

8. Press RD, Kamel-Reid S, Ang D. BCR-ABL1 RT-qPCR for monitoring the molecular response to tyrosine kinase inhibitors in chronic myeloid leukemia. *J Mol Diagn*. 2013;15(5):565–576.

9. Zhen C, Wang YL. Molecular monitoring of chronic myeloid leukemia: international standardization of BCR-ABL1 quantitation. *J Mol Diagn*. 2013;15(5):556–564.

10. Branford S, Fletcher L, Cross NC, et al. Desirable performance characteristics for BCR-ABL measurement on an international reporting scale to allow consistent interpretation of individual patient response and comparison of response rates between clinical trials. *Blood*. 2008;112(8):3330–3338.

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2013-0754-OA by guest on 28 October 2021

11. Muller MC, Cross NC, Erben P, et al. Harmonization of molecular monitoring of CML therapy in Europe. *Leukemia*. 2009;23(11):1957–1963.
12. 1st WHO International Genetic Reference Panel for quantitation of BCR-ABL translocation by RQ-PCR. NIBSC code 09/138. Instructions for use. World Health Organization; 2010.
13. White HE, Matejtschuk P, Rigsby P, et al. Establishment of the first World Health Organization International Genetic Reference Panel for quantitation of BCR-ABL mRNA. *Blood*. 2010;116(22):e111–e117.
14. White HE, Hedges J, Bendit I, et al. Establishment and validation of analytical reference panels for the standardization of quantitative BCR-ABL1 measurements on the international scale. *Clin Chem*. 2013;59(6):938–948.
15. Saldanha J, Silvy M, Beaufils N, et al. Characterization of a reference material for BCR-ABL (M-BCR) mRNA quantitation by real-time amplification assays: towards new standards for gene expression measurements. *Leukemia*. 2007;21(7):1481–1487.
16. Brown JT, Laosinchai-Wolf W, Hedges JB, et al. Establishment of a standardized multiplex assay with the analytical performance required for quantitative measurement of BCR-ABL1 on the international reporting scale. *Blood Cancer J*. 2011;1(3):e13.
17. Beillard E, Pallisgaard N, van der Velden VH, et al. Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR)—a Europe against cancer program. *Leukemia*. 2003;17(12):2474–2486.
18. Gabert J, Beillard E, van der Velden VH, et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia—a Europe Against Cancer program. *Leukemia*. 2003;17(12):2318–2357.
19. Participant summary report: MRD-B minimal residual disease—2012 and 2013 surveys. Northfield, IL: College of American Pathologists.
20. Baccarani M, Cortes J, Pane F, et al. Chronic myeloid leukemia: an update of concepts and management recommendations of European LeukemiaNet. *J Clin Oncol*. 2009;27(35):6041–6051.
21. Cross NC, White HE, Muller MC, Saglio G, Hochhaus A. Standardized definitions of molecular response in chronic myeloid leukemia. *Leukemia*. 2012; 26(10):2172–2175.
22. Kantarjian HM, Baccarani M, Jabbour E, Saglio G, Cortes JE. Second-generation tyrosine kinase inhibitors: the future of frontline CML therapy. *Clin Cancer Res*. 2011;17(7):1674–1683.

Submissions Now Accepted for CAP '15 Abstract Program

Abstract and case study submissions are now being accepted for the College of American Pathologists (CAP) 2015 meeting, which will be held October 4th through the 7th in Nashville, Tennessee. Submissions for the CAP '15 Abstract Program will be accepted from:

Monday, February 9, 2015 through 6 p.m. CT Friday, April 10, 2015

Accepted submissions will be published as a Web-only supplement to the October 2015 issue of the *Archives of Pathology & Laboratory Medicine* and will be posted on the *Archives* Web site. Visit the CAP '15 Web site at www.cap.org/cap15 to access the abstract submission site and additional abstract program information as it becomes available.

Downloaded from http://meridian.allenpress.com/doi/pdf/10.5858/arpa.2013-0754-OA by guest on 28 October 2021