María Sol Ruiz[a], María Belén Sánchez[a], Yuly Masiel Vera Contreras, Evangelina Agrielo, Marta Alonso, María Eugenia Altuna, María Sol Anchordoqui, Mariana Asinari, María Elisa Bonetto, Mauricio Camargo, Isabel Giere, Javier González, Ana Cecilia Granda Alacote, Javier Guerra, Marina Gutiérrez, Cecilia Maldonado, Ricardo Makiya, Gonzalo Manrique, María Eugenia Monaco, Juan Carlos Rozo, Carlos Santamaría, Analía Seravalle, Olga Zea, María Noel Zubillaga, José Mordoh, Irene Larripa and Michele Bianchini*

# Programme for Harmonization to the International Scale in Latin America for *BCR-ABL1* quantification in CML patients: findings and recommendations

## Abstract

**Objectives:** The quantitation of *BCR-ABL1* mRNA is mandatory for chronic myeloid leukemia (CML) patients, and RT-qPCR is the most extensively used method in testing laboratories worldwide. Nevertheless, substantial variation in RT-qPCR results makes inter-laboratory comparability hard. To facilitate inter-laboratory comparative assessment, an international scale (IS) for *BCR-ABL1* was proposed.

**Methods:** The laboratory-specific conversion factor (CF) to the IS can be derived from the World Health Organization (WHO) genetic reference panel; however, this material is limited to the manufacturers to produce and calibrate secondary reference reagents. Therefore, we developed secondary reference calibrators, as lyophilized cellular material, aligned to the IS. Our purpose was both to re-evaluate the CF in 18 previously harmonized laboratories and to propagate the IS to new laboratories.

**Results:** Our field trial including 30 laboratories across Latin America showed that, after correction of raw *BCR-ABL1/ABL1* ratios using CF, the relative mean bias was

[a]María Sol Ruiz and María Belén Sánchez contributed equally to this work.
**\*Corresponding author: Michele Bianchini,** CIO-FUCA, Centro de Investigaciones Oncológicas – Fundación Cáncer, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina; Argenomics, Pilar, Buenos Aires, Argentina; and IMEX, Instituto de Medicina Experimental, CONICET – Academia Nacional de Medicina, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina, E-mail: mbianchini@conicet.gov.ar
**María Sol Ruiz, María Belén Sánchez, Yuly Masiel Vera Contreras and José Mordoh:** CIO-FUCA, Centro de Investigaciones Oncológicas – Fundación Cáncer, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina
**Evangelina Agrielo:** LEB, Laboratorio de Especialidades Bioquímicas, Bahia Blanca, Buenos Aires, Argentina
**Marta Alonso:** Hospital Nac. A. Posadas, El Palomar, Buenos Aires, Argentina
**María Eugenia Altuna:** Clinica Dr. Roberto Raña, Neuquen, Argentina
**María Sol Anchordoqui:** Argenomics, Pilar, Buenos Aires, Argentina
**Mariana Asinari:** Hospital Privado Universitario, Cordoba, Argentina
**María Elisa Bonetto:** Hospital Dr. G. Rawson, San Juan, Argentina
**Mauricio Camargo and Olga Zea:** Genética Lab, Medellín, Antioquia, Colombia

**Isabel Giere:** Fundaleu, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina
**Javier González:** Meyer Lab, Asunción, Paraguay
**Ana Cecilia Granda Alacote:** Laboratorios Medicos A Yuen y G Rios, Lima, Peru
**Javier Guerra:** Nanopharmacia Diagnostica, Ciudad de Mexico, Mexico
**Marina Gutiérrez:** Stamboulian, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina
**Cecilia Maldonado:** ManLab, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina
**Ricardo Makiya:** FIBIO, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina
**Gonzalo Manrique and María Noel Zubillaga:** ASESP, Asociación Española, Montevideo, Uruguay
**María Eugenia Monaco:** Laboratorio Tucuman, San Miguel de Tucumán, Tucumán, Argentina
**Juan Carlos Rozo:** UDHO, Unidad de Diagnóstico Hemato Oncológico, Cali, Valle del Cauca, Colombia
**Carlos Santamaría:** Hospital de Niños Carlos Sáenz Herrera, San José, Costa Rica
**Analía Seravalle:** CIBIC, Rosario, Santa Fe, Argentina
**Irene Larripa:** IMEX, Instituto de Medicina Experimental, CONICET – Academia Nacional de Medicina, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina

significantly reduced. We also performed a follow-up of participating laboratories by annually revalidating the process; our results support the need for continuous revalidation of CFs. All participating laboratories also received a calibrator to determine the limit of quantification (LOQ); 90% of them could reproducibly detect *BCR-ABL1*, indicating that these laboratories can report a consistent deep molecular response. In addition, aiming to investigate the variability of *BCR-ABL1* measurements across different RNA inputs, we calculated PCR efficiency for each individual assay by using different amounts of RNA.

**Conclusions:** In conclusion, for the first time in Latin America, we have successfully organized a harmonization platform for *BCR-ABL1* measurement that could be of immediate clinical benefit for monitoring the molecular response of patients in low-resource regions.

**Keywords:** *BCR-ABL1*; harmonization; Ph-IS; RT-qPCR.

## Introduction

Residual disease in chronic myeloid leukemia (CML) patients undergoing therapy with tyrosine kinase inhibitors (TKIs) is measured by assessing the quantity of transcripts of the *BCR-ABL1* fusion gene in peripheral white blood cells. This analysis is carried out with molecular tests, based on quantitative reverse-transcription PCR (RT-qPCR) technology; however, the wide array of preanalytical and analytical methods used worldwide and the absence of consensus guidelines have led to large variations in quantitative *BCR-ABL1* measurements, which hamper inter-laboratory comparative studies, patient manageability and standardized definition of treatment response [1]. 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. It is now recognized that monitoring *BCR-ABL1* ratios on the so-called international scale (IS) is vital in the management of patients with CML and for improving outcomes. Efforts to harmonize procedures to measure *BCR-ABL1* fusion transcripts have included important investments in sample exchange programs to derive laboratory-specific conversion factors (CFs); these efforts showed improvements in inter-laboratory concordance rates, but the process is laborious and limited due to the lack of a common set of reference samples that can be shared on a global scale [2, 3]. This requirement was addressed in part by the formulation and validation of the first World Health Organization

(WHO) International Genetic Reference Panel for quantitation of *BCR-ABL1* by RT-qPCR [4]. The WHO primary standards consist of a four-level panel of e14a2-positive lyophilized cell line dilutions. Each level has an assigned IS value, which was obtained by repeated testing of each sample level in expert IS-standardized laboratories. Unfortunately, the stock of the WHO primary standards is limited, and their accessibility has been restricted to manufacturers of testing kits or secondary reference standards. In principle, these secondary reagents might be formulated in different ways, such as freeze-dried cells, cell lysates or synthetic RNAs, but they would need to be in a form that enables calibration to the IS (i.e. reagents for which a *BCR-ABL1*/control gene ratio can be derived). Accordingly, as recently reported by us [5], we developed secondary reference materials as lyophilized cellular calibrators, aligned to the IS percent ratios through the WHO primary standards. Also, we piloted an international, multicenter, harmonization program (Ph-IS) by using five-level cellular calibrator samples. Based on this experience, the first in Latin America, in this work we report an update at 3 years of the program and several suggestions in order to improve accuracy in the quantification of *BCR-ABL1*.

## Materials and methods

### Participant laboratories

The study was conducted by a single reference laboratory, Genetic and Hematologic Laboratory (from here on called RefLab), belonging to the IMEX (CONICET-National Academy of Medicine, Buenos Aires, Argentina). Thirty laboratories were invited to take part in the international collaborative program and were chosen because they routinely perform *BCR-ABL1* testing. These laboratories form the *LABS* (*L*atin *A*merica *B*CR-ABL1 *S*tandardization) group. Nineteen laboratories were from Argentina, while the remaining 11 laboratories were from Mexico, Uruguay, Paraguay, Costa Rica, Panamá, Peru, Colombia, Guatemala and Ecuador (Supplementary Table 1). The materials were analyzed in each laboratory according to its own RNA extraction and RT-qPCR protocol. The following data were recorded: date of RNA extraction, total micrograms of RNA, $A_{260}/A_{280}$, $A_{260}/A_{230}$, *BCR-ABL1* transcript value (cycle threshold [$C_t$] value and copy number) *ABL1* transcript value ($C_t$ value and copy number), *BCR-ABL1/ABL1* ratio before conversion to IS (*BCR-ABL1IS* will be used to indicate BCR-ABL1/ABL1 values that have been corrected using the CF), and slope and efficiency of standard curves. Standard curves of known amounts of *BCR-ABL1* and *ABL1* allow calculation of the %ratio in the test sample. Standard curves were obtained by using three different types of plasmid dilutions depending on the RT-qPCR method employed in each laboratory; of the 31 laboratories (including RefLab), 15 of them used Molecular MD Kit (MolecularMD, Cambridge, MA, USA), which includes a unique six-plasmid dilution

series (from $3\times10^5$ to $3\times10^0$) for both *BCR-ABL1* and *ABL1* targets. Twelve laboratories used Ipsogen plamids (Qiagen, Germantown, MD, USA) – four of them from the Ipsogen® BCR-ABL1 Mbcr IS-MMR DX kit, which provides a unique six-plasmid dilution series (from $1\times10^6$ to $1\times10^1$) for both *BCR-ABL1* and *ABL1* targets, while the others eight laboratories, which employed an LDT approach, used Ipsogen *ABL* Control Gene (four Standards [from $1\times10^3$ to $1\times10^6$]) and Ipsogen *BCR-ABL1* Mbcr Standards (5 standards [from $1\times10^1$ to $1\times10^6$]). The remaining four laboratories used the ERM® Certified Reference Material from European Commission – Joint Research Centre Institute for Reference Materials and Measurements (IRMM), which provides the six ERM-AD623 plasmid solutions (from $1.08\times10^6$ to $1\times10^1$). Raw data generated by the 30 participating laboratories were directly sent by e-mail from each site to RefLab, by using a specific form.

## Reference laboratory validation

RefLab employed a one-step RT-qPCR approach based on the MolecularMD kit (MRDx® BCR-ABL Test, NY, USA) and a Rotor-Gene® Q MDx 5plex HRM (high-resolution melting) platform (Qiagen, Stockach, Germany). RefLab obtained the WHO primary standards (NIBSC code 09/138) from the United Kingdom National Institute for Biological Standards and Control (Potters Bar, United Kingdom). In order to ensure the consistent performance of the entire process along time, two different quality control (QC) RNA samples with a high and low *BCR-ABL1* level were processed in the same way as the patient samples in every run. QC samples had pre-established *BCR-ABL1* values and, for the run to be accepted, the values had to be within a defined range, which was based on the standard deviation (SD) of the analytical system. The yearly mean QC values since 2010 are detailed in Supplementary Table 2.

## Secondary calibrator production

A total of 20 batches of cellular calibrators were independently prepared according to standard operating procedures, in order to satisfy the need for calibrators, in the Latin American region. 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. Cell lines were grown in RPMI 1640 medium (Gibco) supplemented with 10% (for K562) or 20% (for HL60) fetal bovine serum (Natocor, Cordoba, Argentina). Formulations were planned to target IS% ratios close to each of the established TKI clinical response criteria, that is, 10% (complete hematologic response), 1% (complete cytogenetic response), 0.1% (MR[3.0], major molecular response), and 0.01% and 0.001% (MR[4.0] and MR[5.0] respectively, deep molecular response). The last dilution is expected to be close to the limit of detection for most of RT-qPCR methods. Samples were stabilized by lyophilization [5], in glass ampules filled with $2\times10^6$ cells (1 mL) (AdVantage-Plus freeze dryer, VIRTIS). Freeze-dried material was stored at –20 °C (range between –18 ° and –22 °C) and shipped at room temperature between 15 ° and 25 °C. Each panel of calibrators (i.e. CAL-IS, *Cal*ibrator to *I*nternational *S*cale) was labeled in a blinded fashion such that participating laboratories received 2 ampules from each of the five *BCR-ABL1* expression levels (named CAL-IS_01 to CAL-IS_05), thus representing in total 10 blinded ampules per panel. For evaluation of the homogeneity and stability, before dispatching, we checked that each lot met all predefined QC criteria for RNA yield, purity and minimum ABL1 copy number per PCR reaction (see Supplementary Materials and Methods). For the assignment of %IS reference values, each batch was evaluated in quintuplicate (from two randomly picked ampoules from each level of *BCR-ABL1* expression) at the RefLab (see Supplementary Materials and Methods).

## Peripheral blood samples

Peripheral blood samples were collected from two CML patients at diagnosis (Dx) before any treatment. The diagnosis of CML was determined through the detection of the Philadelphia chromosome using cytogenetic evaluation and end-point RT-PCR for *BCR-ABL1*. In each laboratory, RNA was extracted using Trizol (ThermoFisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. Patients provided written informed consent, and the study was approved by the local research Ethics Committee.

## RNA input curve

For the evaluation of efficiency of the RT-qPCR assay for *BCR-ABL1* and *ABL1*, each laboratory performed a curve with different amounts of RNA obtained from CAL-IS_03 (MR[3.0]), using 125 ng, 250 ng, 500 ng and 1000 ng of RNA input per RT-qPCR reaction; three replicates were run per sample at each input level. For evaluation of the efficiency of the method using samples from CML patients at Dx, each laboratory performed a curve with 250, 500 and 1000 ng of RNA input per RT-qPCR reaction; two replicates were run at each input level.

## Data analysis

Laboratory-specific CFs were calculated as the antilog of the arithmetic mean of the differences between log-transformed reference IS% ratios and log-transformed raw percentage ratios measured in each laboratory for each panel and level. All data analyses assumed a normal-like distribution of results after $Log_{10}$ transformation of the percent ratios. Bias and correlation analyses were performed by using the Bland-Altman method and least squares linear regression analyses. Variability at individual sample levels, across multiple sample levels and in pooled analyses was assessed by using $Log_{10}$-transformed percentage ratios before calculation of the 95% limits of agreement (LOA). To determine the LOA of the RefLab method (intrinsic method variation), RNA samples were analyzed in replicates within the reference laboratory encompassing variables such as the operator, day of analysis, reagent batches, and calibration status of the instrument and pipettes. The spread of results as estimated by the 95% LOA was plus or minus 2.5-fold of the mean, which is an indicator of within-method variability. For each method, the limit of quantification (LOQ) was calculated by considering the upper level of the SD interval from the mean. The $CF_{Index}$ was calculated as the ratio between CFs derived from two consecutive rounds of calibration ($CF_n/CF_{n+1}$).

# Results

## LABS group status in 2019

The whole study design comprises five principal steps as illustrated in Supplementary Figure 1 and previously reported by us [5]. As detailed in the Materials and method section, two panels of CAL-IS were sent to each one of the 30 clinical laboratories for evaluation (Step 4, Supplementary Figure 1). Each laboratory tested the panel with unique *BCR-ABL1* assays, resulting in a total of 30 *BCR-ABL1* tests included in this report. Only Trizol RNA extraction method was used by following manufacturer indications; for 16 laboratories that employed a two-step approach for cDNA synthesis, two RT reactions were performed for each of the extracted RNA samples on two different days, giving a total of 20 cDNA samples. Samples for *BCR-ABL1* and the control gene were analyzed using the established RT-qPCR methods of the participating laboratories. Six different types of RT-qPCR machines were used (Supplementary Table 1). Twelve of the participants used laboratory-developed test as their RT-qPCR protocol based on Gabert et al. [6] and Beillard et al. [7], while the remaining 18 laboratories used commercial kits (Supplementary Table 1).

The laboratories were asked to conduct two studies with the secondary panel. In Study 1, in order to estimate the specific CF to the IS and to assess the performance of the *BCR-ABL1* tests, each laboratory quantified CAL-IS_01 to CAL-IS_04 on four different days, carrying out a total of 72 RT-qPCR reactions. In Study 2, in order to determine the LOQ, 2 ampoules of CAL-IS_05 were tested on four different days, carrying out a total of eight reactions.

Raw *BCR-ABL1/ABL1* ratios generated in each laboratory were used to calculate the specific CF to IS (Step 5, Supplementary Figure 1), as detailed in Supplementary materials and methods, and to assess the sensitivity, linearity and precision of local RT-qPCR methods (Table 1). Raw percentage ratios were consistent with a 10-fold serial dilution and were linear for the first four levels. All laboratories obtained *ABL1* copy numbers above the minimum copy number requirements (>32,000 *ABL1* copies) for almost all samples evaluated (less than 3% of tests were rejected) when considering all runs performed. Inter-laboratory variability and IS standardization were assessed relative to nominal IS% ratios independently assigned to each CAL-IS by direct comparison to the WHO primary standards. Before harmonization, the relative mean bias ranged from −0.41-fold to +0.40-fold in the 30 laboratories, indicating a relatively wide inter-laboratory variability (Figure 1A). Following conversion of raw percentage ratios to IS% ratios using laboratory-specific CF, the relative mean bias ranged from −0.027-fold to +0.033-fold (Figure 1B), indicating a significant reduction in inter-laboratory variability. The intra-laboratory precision was good with an average 95% LOA of 2.1-fold and less than 2.5-fold in most cases, with the corresponding level-specific mean CV between 6% and 45% (mean = 21%, with 26 out of 30 laboratories with a mean CV ≤30%). In Table 1, we also report the last CF values for each laboratory with the corresponding 95% LOA, the median and the range. In total, 75 CFs were derived if considering all laboratories.

To survey whether changes in the detection system, methodology, reagents or operator could cause a shift in the CF value, we performed a follow-up of participating laboratories by annually revalidating the process. As shown in Figure 2, without considering RefLab, 21 out of 30 laboratories could repeat the entire process of harmonization at least once. Of the 21 laboratories with complete data, nine had stable CFs (defined as a $CF_{index}$ within 0.75–1.33, green points in the graph; Figure 2), in all rounds of validation. Three laboratories improved their stability since the first $CF_{index}$ calculation (a mix of red and green points over time); nine laboratories were outside the optimal range in all rounds of validations. This inconsistency is also reflected by the median and range columns reported in Table 1; particularly, Lab #21 shows an outlier value due to simultaneous modification of the platform and methodology. These results support the need for continuous revalidation of CFs.

All participating laboratories received the fifth calibrator (CAL-IS_05); as expected, *BCR-ABL1* was inconsistently detected below the linear quantitative detection range of the RT-qPCR method in the lowest positive sample. However, CAL-IS_05 is useful to calculate the LOQ for each method; as indicated in the Materials and methods section, LOQ was calculated by considering the upper level of the SD interval from the mean (Table 1). The overall *BCR-ABL1* detection rate ranged from 0% (1 out of 30 laboratories) to 100% (23 out of 30 laboratories). In total, 27 out of 30 laboratories (90%) could reproducibly detect *BCR-ABL1* in CAL-IS_05; these laboratories can report a consistent MR[4,5].

## Evaluation of reverse transcription efficiency for target and reference genes

With the aim to investigate the variability of *BCR-ABL1* measurements across different RNA inputs, we calculated the PCR efficiency (Ef) for each individual *BCR-ABL1* and

**Table 1:** For each local RT-qPCR method, the specific conversion factor, reproducibility, precision and sensitivity are reported.

| Laboratory number | Last CF | 95% LOA | Median | Range | LOQ | MR4.5 | CF index | CV, % | Log reduction |
|---|---|---|---|---|---|---|---|---|---|
| Ref lab | 0.75 | 0.44–1.32 | 0.7 | 0.68–0.75 | 0.0003 | Yes | 0.89 | 19 | 5.5 |
| Lab#1 | 0.84 | 0.45–1.57 | NA | NA | 0.0005 | Yes | NA | 27 | 5.3 |
| Lab#2 | 0.53 | 0.34–0.81 | 0.53 | 0.49–0.72 | 0.0018 | Yes | 1.36 | 18 | 4.7 |
| Lab#3 | 0.94 | 0.78–1.14 | 0.83 | 0.62–0.94 | 0.0007 | Yes | 0.88 | 11 | 5.2 |
| Lab#4 | 1.10 | 0.43–2.71 | 1.1 | 1.01–1.10 | 0.0018 | Yes | 1.0 | 35 | 4.7 |
| Lab#5 | 0.76 | 0.50–0.78 | 0.74 | 0.63–0.76 | 0.0021 | Yes | 0.95 | 17 | 4.7 |
| Lab#6 | 0.81 | 0.69–1.49 | 1.01 | 0.81–1.2 | 0.0017 | Yes | 1.48 | 23 | 4.8 |
| Lab#7 | 0.45 | 0.12–1.65 | NA | NA | 0.0030 | Yes | NA | 25 | 4.5 |
| Lab#8 | 1.72 | 0.89–3.37 | NA | NA | UND | No | NA | 33 | NA |
| Lab#9 | 0.71 | 0.52–0.84 | 0.69 | 0.67–0.71 | 0.00085 | Yes | 0.97 | 9 | 5.1 |
| Lab#10 | 0.73 | 0.59–0.97 | 0.72 | 0.71–0.73 | 0.0017 | Yes | 0.97 | 13 | 4.8 |
| Lab#11 | 0.68 | 0.23–2.00 | 1.02 | 0.68–1.35 | 0.0004 | Yes | 1.98 | 45 | 5.4 |
| Lab#12 | 0.40 | 0.35–0.44 | 0.72 | 0.40–1.10 | 0.0031 | Yes | 1.90 | 7 | 4.5 |
| Lab#13 | 0.62 | 0.41–0.69 | 0.54 | 0.39–0.62 | UND | No | 0.76 | 24 | NA |
| Lab#14 | 2.51 | 1.51–4.17 | 1.16 | 0.52–2.51 | UND | No | 0.46 | 32 | NA |
| Lab#15 | 0.69 | 0.34–1.42 | NA | NA | 0.0030 | Yes | NA | 16 | 4.5 |
| Lab#16 | 0.59 | 0.32–1.00 | 0.83 | 0.56–1.09 | 0.0019 | Yes | 1.95 | 22 | 4.7 |
| Lab#17 | 1.21 | 0.88–1.59 | NA | NA | 0.0010 | Yes | NA | 13 | 5.0 |
| Lab#18 | 0.57 | 0.35–0.91 | 0.83 | 0.57–1.09 | 0.0012 | Yes | 1.91 | 18 | 4.9 |
| Lab#19 | 0.59 | 0.37–0.95 | NA | NA | 0.0014 | Yes | NA | 23 | 4.9 |
| Lab#20 | 0.78 | 0.51–1.18 | NA | NA | 0.0012 | Yes | NA | 30 | 4.9 |
| Lab#21 | 0.73 | 0.46–1.17 | 0.73 | 0.35–2.35 | 0.0016 | Yes | 0.48 | 21 | 4.8 |
| Lab#22 | 0.59 | 0.50–0.68 | 0.93 | 0.59–1.58 | 0.0010 | Yes | 0.40 | 6 | 5.0 |
| Lab#23 | 0.89 | 0.66–1.19 | 1.15 | 0.90–1.17 | 0.0010 | Yes | 1.02 | 19 | 5.0 |
| Lab#24 | 1.10 | 0.79–1.73 | 1.1 | 0.98–1.33 | 0.0008 | Yes | 1.21 | 20 | 5.1 |
| Lab#25 | 0.89 | 0.62–2.97 | 1.13 | 0.89–1.36 | 0.0012 | Yes | 0.65 | 38 | 4.9 |
| Lab#26 | 0.85 | 0.50–1.46 | NA | NA | 0.0019 | Yes | NA | 23 | 4.7 |
| Lab#27 | 0.61 | 0.42–0.80 | 1.22 | 0.61–1.82 | 0.0010 | Yes | 2.98 | 12 | 5.0 |
| Lab#28 | 0.75 | 0.57–0.99 | 0.75 | 0.62–0.75 | 0.0014 | Yes | 1.0 | 17 | 4.9 |
| Lab#29 | 0.95 | 0.82–1.66 | 1.53 | 0.79–1.55 | 0.00055 | Yes | 1.0 | 24 | 5.3 |
| Lab#30 | 1.48 | 1.04–2.10 | NA | NA | 0.0013 | Yes | NA | 14 | 4.9 |

CF index, values between 1.33 and 0.75 are considered stable. CF, conversion factor; CV, coefficient of variation; LOA, limit of agreement; log reduction, maximum measurable log reduction; LOQ, limit of quantification; MR4.5, achievable (yes) or not achievable (no) 4.5 molecular response; NA, not available; UND, undeterminable.

reference gene assay. Accordingly, we performed a pilot study including eight laboratories (seven from Argentina and one from Uruguay) that are part of the LABS network. Each laboratory performed quantification of *BCR-ABL1* and *ABL1* using four different amounts of RNA during the RT step (1000, 500, 250 and 125 ng). Total RNA was obtained from secondary calibrator samples and processed using the routine method in each laboratory. Evaluation of Ct vs. Log (RNA) for each gene gave different results depending on the laboratory, which could be grouped into three categories: (i) four out of eight laboratories (50%) showed comparable efficiencies for *BCR-ABL1* and *ABL1* genes, with values close to the ideal Ef=100% (RefLab, Labs #3, #11, #12); (ii) three out of eight laboratories (38%) showed comparable efficiencies for *BCR-ABL1* and *ABL1* genes, with values deviated from the ideal Ef (Labs #2, #4, #21); and (iii) one out of eight laboratories (15%) showed

efficiencies for *BCR-ABL1* and *ABL1* genes that differed significantly (Lab #16, $p < 0.05$) (Figure 3).

## BCR-ABL1/ABL1 ratio at diagnosis

The *BCR-ABL1IS* transcript rate decline from baseline to 3 months could be more predictive than 10% IS level at 3 months. However, discordant results have been reported about the feasibility of *ABL1* as an internal gene control, as it is not reliable for *BCR-ABL1* quantification above 10% [8]. Given that all laboratories included in the harmonization program used *ABL1* as the reference gene, we evaluated how different RNA inputs (1000, 500 and 250 ng) affected the ratios using samples from CML patients at Dx. In order to minimize variability, quantified aliquots from the same CML Dx sample were
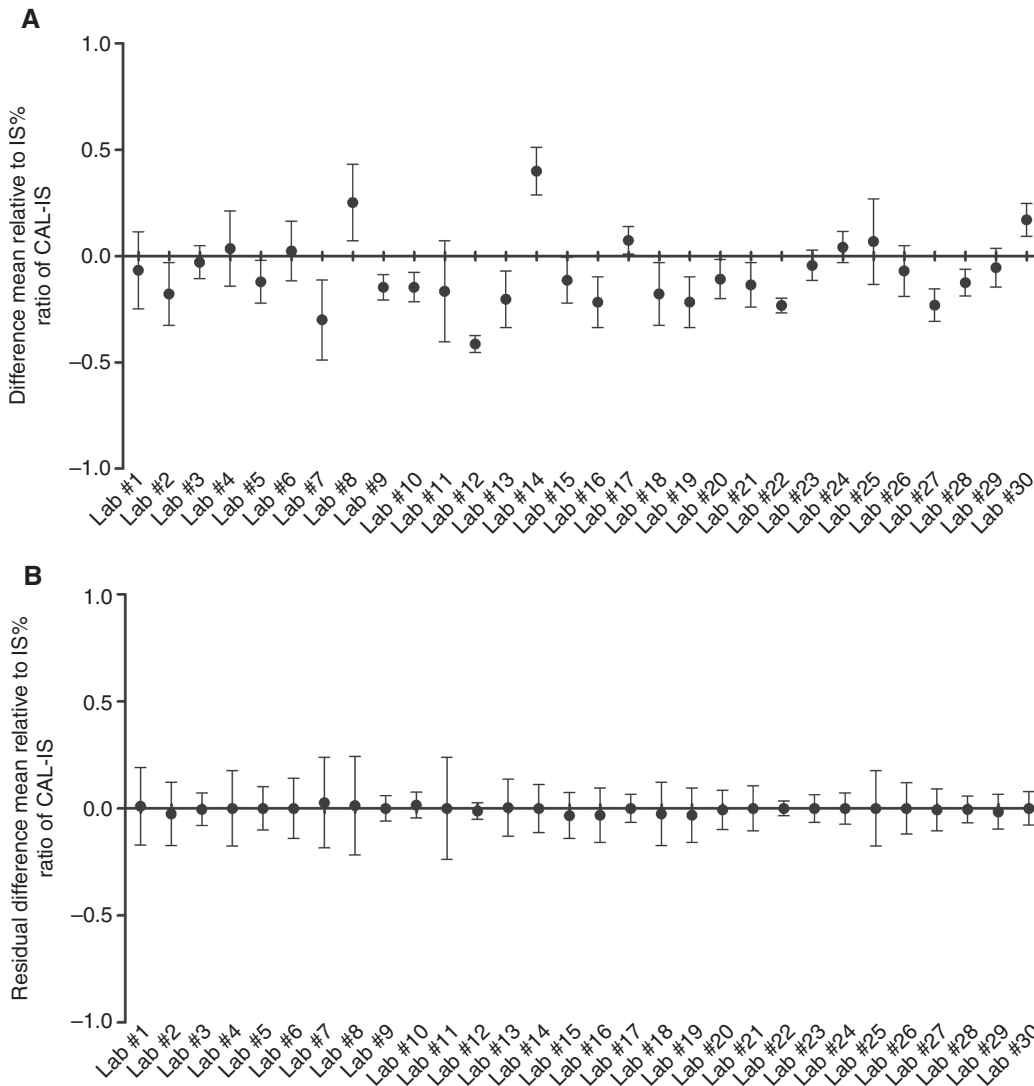
**Figure 1:** Graphs showing relative mean bias for each of the 30 laboratories.
(A) Relative difference between nominal and measured IS% ratios for all calibrator levels combined (black circle). (B) Residual difference after CF-conversion. Error bars show 95% LOA interval.

distributed to the participating laboratories. Only six laboratories participated at this stage. Despite being harmonized to the IS, we observed significant variation in the *BCR-ABL1IS* ratios among participating laboratories (CV%: 74, 74 and 85 for 250, 500 and 1000 ng, Figure 4), ranging from 0.2-fold to 2.6-fold variation from the mean value of all laboratories. This observation is in accordance with the limit of applicability of the CF, which should be used only for ratios included in the range of secondary calibrators (around 10%–0.01%). Moreover, comparing *ABL1* and *BCR-ABL1* $C_t$ against RNA input, we observed that some laboratories showed different performance for both genes (Figure 5). In order to elucidate which factors, present at Dx samples but absent in secondary calibrators, could affect qPCR performance, we

evaluated the effect of removing possible inhibitors and residual DNA. We observed differences neither between pre- and post-purification, nor between pre- and post-treatment with DNAse (Supplementary Figure 2). We further evaluated the effect of varying concentrations of primers and probes in the qPCR reaction. We observed higher variation in *BCR-ABL1IS* ratios with higher RNA input (Supplementary Figure 3).

## Discussion

Despite efforts to establish standardized protocols for *BCR-ABL1* fusion transcript quantification, there is still substantial variation in the way through which RT-qPCR
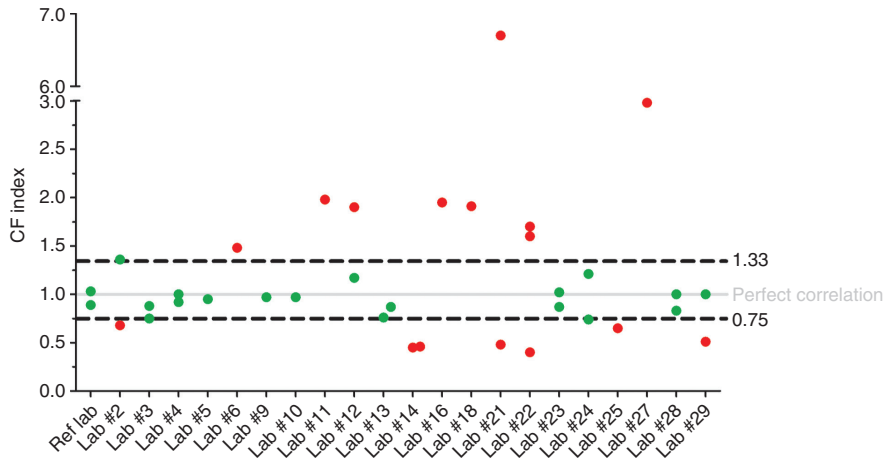
**Figure 2:** Graph showing the follow-up of 21 participating laboratories.
The CF index represents the ratio of two CFs obtained in two consecutive rounds of calibration. The gray line indicates a perfect match between consecutive CFs; CF indexes within 0.75–1.33 (optimal range) are considered stable (green points in the graph). Laboratories outside the optimal range are considered unstable.

is carried out. In this report, we intentionally used the term "harmonization" to define the entire program, to mark a difference with respect to the concept of "standardization"; standardization is about adopting a uniform approach to tackle specific work; on the other hand, harmonization has a broader scope than standardization, as it seeks equivalence of measurement results among different routine measurement procedures over time (i.e. specimen collection and handling, preanalytical procedures, harmonizing measurement results, interpretation of results in medical contexts). Harmonization should allow better comparability between laboratories and lead to the standardization of patient assessment.

Despite the fact that commercial kits commonly used for RT-qPCR analysis of *BCR-ABL1* are supplied with a CF value or a unique IS-calibrator in order to correct the results to the IS, it is important to consider at least two drawbacks, when a "portable" CF is used: first, when using a preestablished CF, pre-analytical (e.g. white blood cell lysis, RNA extraction and purification) and analytical (e.g. PCR platform) steps are not included in the IS calibration process. In our experience, such factors are critical and should be always evaluated during the calibration procedure. The second important weakness of using "portable" CF derives from the fact that IS standardization cannot be achieved with a CF value derived from a single reference sample; in our opinion, both precision (intra- and inter-laboratory reproducibility) and accuracy (mean bias relative to the IS) should be assessed across a clinically appropriate interval of IS% ratios. For this reason, we established and validated cellular calibrators produced by serial dilution of the K562 Ph-positive cell line in the

HL-60 Ph-negative cell line, following the same scheme used for WHO primary standards. The establishment and validation of the WHO primary standards and secondary reagents was certainly an important milestone for IS harmonization around the world [4, 9]. In the present study, we show that secondary reference biological calibrators anchored to the WHO primary standards can decrease inter-laboratory variability. Log *BCR-ABL1/ABL1* values for the same calibrators were compared between RefLab and local laboratories applying the Bland-Altman bias plot. 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. Accordingly, by comparing the absolute value of kit-derived CF vs. CAL-IS-derived CF in most cases, we observed a significant difference between the two values, reinforcing the concept that harmonization by secondary calibrators is absolutely necessary.

In addition to investigate the effectiveness of this process, a second aim of this program was to examine the stability of CFs over time. CF value stability within each laboratory is a very important parameter to ensure comparability of results along time. In order to judge the stability of each laboratory's methodology, a CF index (ratio of two consecutive rounds of calibration) was calculated and evaluated according to its capability to achieve optimum concordance of results. In most cases, we obtained concordant CFs, in particular when no methodology modifications were introduced since the time of the last calibration. However, nine laboratories did not maintain their CFs over
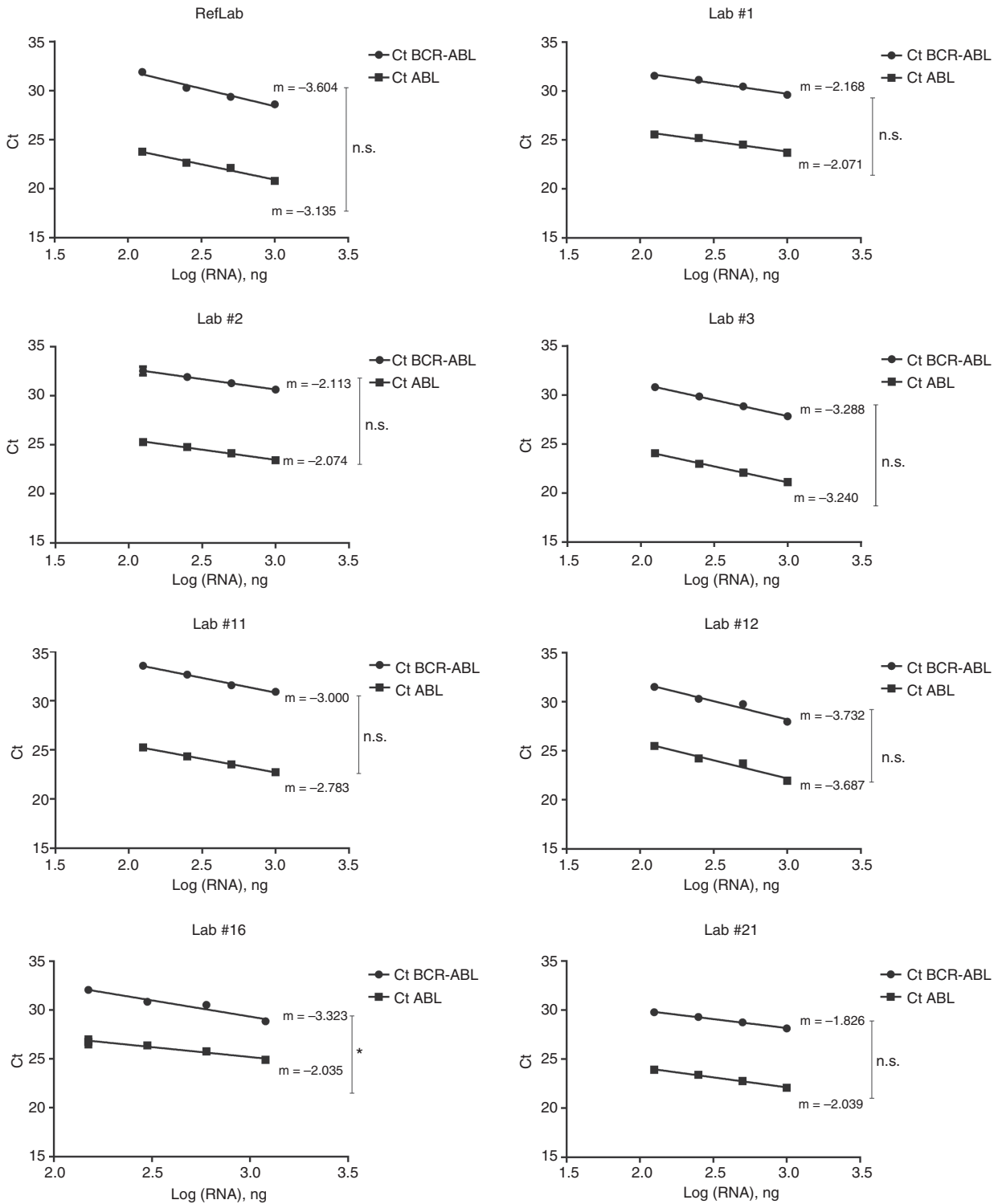
**Figure 3:** Graphs showing the correlation between the threshold cycle (Ct) number and RNA input in log scale (Log RNA). m, slope of the line; n.s., no significant difference between slopes.

time and only in six of them this could be attributed to modifications of reagents or equipment. These data indicate that CFs may be unstable in some laboratories even in the absence of significant changes to laboratory protocols.

According to Branford et al. [8, 10], minor alterations to analytical systems may have significant impacts on the measurements, requiring CF recalculation. Indeed, imperceptible changes are always occurring in laboratories,
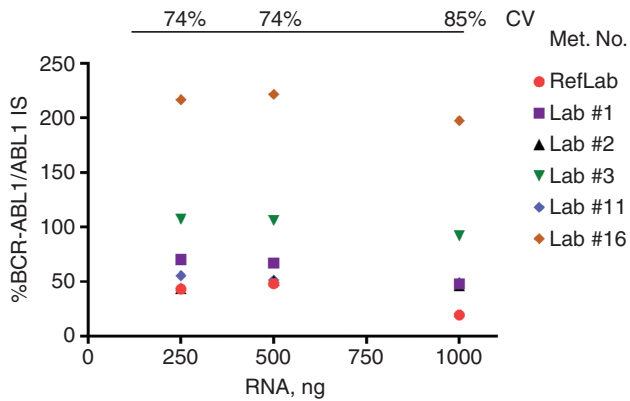
**Figure 4:** Graph showing *BCR-ABL1/ABL1* ratios at diagnosis for six methods (Met. No.) using *ABL1* as an internal control gene. Quantitative PCRs were run on five different platforms (RotorGene for RefLab, ABI7500 for LAB#1 and LAB#2, Cobas Z480 for LAB#3, Abbott 2000 for LAB#11 and LightCycler2.0 for LAB#16). Four laboratories (RefLab, LAB#3, LAB#11 and LAB#16) employed the Molecular MD kit and the included plasmid dilution series for standard curve construction; on the other hand, LAB#1 and LAB#2 employed an LDT approach, with ERM-AD623 or Ipsogen plasmid curve, respectively. CV, coefficient of variation.

such as instrument aging, reagent lot changes, operator differences, etc., which might affect the stability of the measurements. Therefore, the stability of CF in IS laboratories should be monitored regularly. In daily work, continuous CF monitoring can be carried out by including internal QC samples in every run to determine whether the CF needs to be validated or not. For those laboratories with unstable CFs, we suggest revalidation and a rigorous check-up of their internal procedure to identify potential sources of variation, whereas those with stable CFs should be assessed on a yearly basis. Finally, we strongly recommend to revalidate the CF every time the procedure is modified like introduction of a new instrument or important methodology changes which might affect the stability of the measurements.

RT-qPCR assays aim to measure reliably to at least $MR^{4.5}$, depending on the quality of the sample received for testing; a deep molecular response is defined as *BCR-ABL1IS* values below 0.01% IS ($MR^{4.0}$) where $MR^{4.5}$ is <0.0032% IS, and $MR^{5.0}$ <0.001%. *BCR-ABL1* detection limits of $MR^{4.0}$ to $MR^{5.0}$ are adequate to meet the clinical requirements, as there is not any evidence that the detection of *BCR-ABL1* below 0.001% IS could indicate any clinical benefit in terms of survival advantage or a difference in the rate of failure events [11]. However, molecular sensitivity, precision and accuracy are extremely important parameters for TKI cessation studies. The CAL-IS_05 sample was used to validate the sensitivity and $MR^{4.5}$ detection

capability of the assays. Most of the methods successfully detected *BCR-ABL1* in the CAL-IS_05, demonstrating the capacity to quantitate extremely small amounts of *BCR-ABL1* mRNA in patient samples. However, three laboratories did not reach this level of sensitivity; in these cases, we suggested PCR optimization guidelines to improve assay performance.

In order to perform accurate quantification, target and reference genes should perform similarly across different conditions. Optimization of quantitative PCR requires the evaluation of amplification efficiency by means of a standard curve for each gene. These curves are usually performed with DNA, such as commercial or in-house-developed plasmids. However, the retrotranscription (RT) step can also add significant variation because of differential efficiency of RT for each gene, the presence of inhibitors of the RT process and/or qPCR in the sample, and the effect of different RNA inputs during the RT step [12]. Therefore, evaluation of global efficiency using RNA samples is highly desirable.

*BCR-ABL1IS* ratio is intended for CML samples from patients under treatment. Recent reports describing the impact of early reduction of *BCR-ABL1* transcripts from baseline suggest that quantification of *BCR-ABL1* transcripts on Dx samples could be of clinical utility [13, 14]. Nevertheless, for high levels of leukemic burden, results are less reliable because of the lack of linearity particularly for laboratories using *ABL1* or *GUSB*, as an internal control gene [15]. Results obtained here discourage the use of *BCR-ABL1* transcript rate decline from baseline, while using *ABL1* as the control gene, because *BCR-ABL1* quantification is not reliable above 10% IS when data are obtained from different laboratories, despite them belonging to a harmonization program. In order to reduce variability of diagnostic ratios, a lower RNA input (<300 ng) resulted in lower variability when altering qPCR conditions, suggesting that higher inputs of RNA, where larger amounts of leukemic transcripts are present, may imply a condition where reagents are limiting factors. Alternatively, as recently reported by Dulucq et al. [16], *BCR-ABL1* transcript decrease from month 3 to month 6 using *ABL1* as an internal control gene can efficiently identify patients (>1% IS at 6 months) at higher risk of event.

In conclusion, for the first time in Latin America, we have successfully organized a harmonization platform for *BCR-ABL1* measurement (www.ph-is-com). We identified existing flaws in *BCR-ABL1* quantification and suggested several recommendations to improve comparability between laboratories. We believe that this platform can improve the reproducibility and accuracy
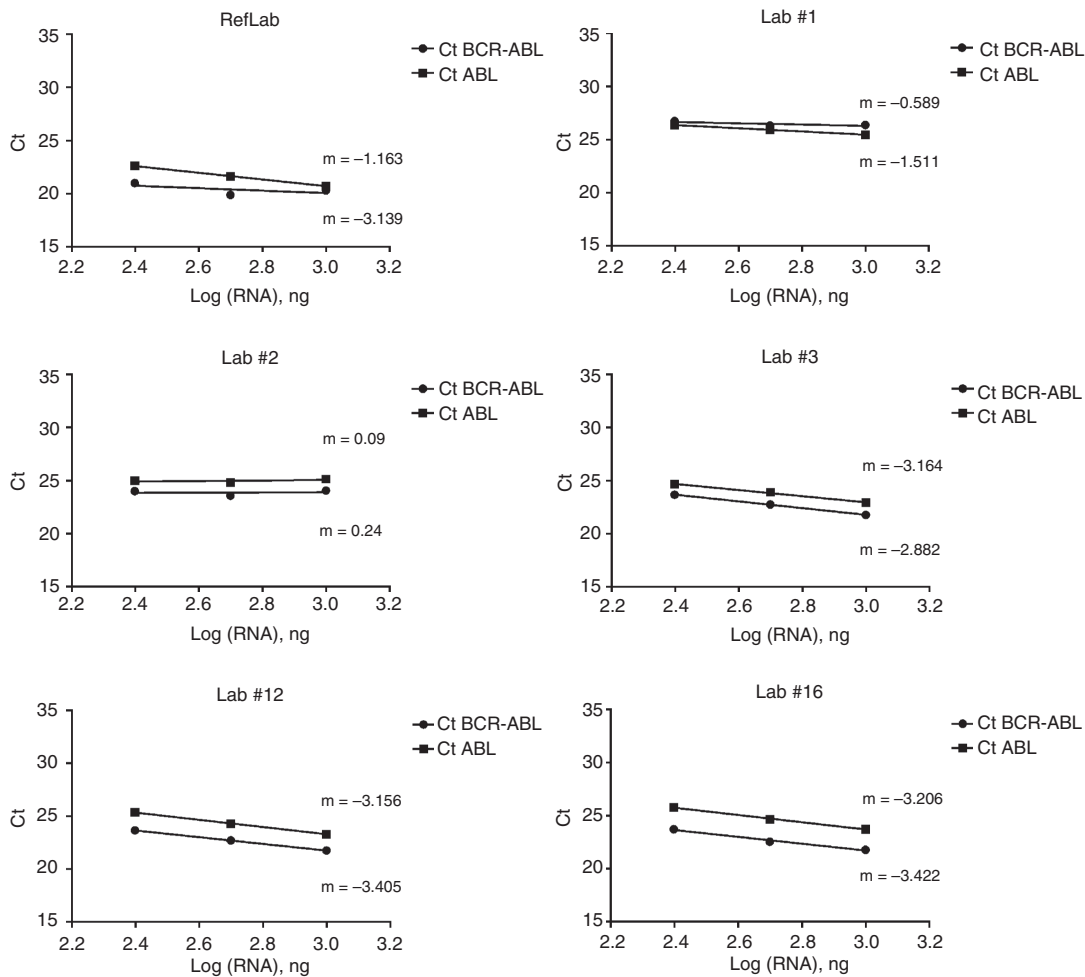
**Figure 5:** Graphs showing the correlation between the threshold cycle (Ct) number and RNA input in log scale (Log RNA) for a Dx sample. m, slope of the line.

of *BCR-ABL1* detection to ensure robust laboratory diagnostic capacities in Latin America.

**Author contributions:** M.B. and I.L. designed the program of harmonization. M.S.R. and M.B.S performed most of the experiments. M.S.R. and M.B. performed data analysis; all the authors contributed with technical support in RT-qPCR runs, reviewed the data, drafted parts of the manuscript and participated with helpful discussion. M.B. wrote the manuscript and supervised the entire work. All authors read and approved the final manuscript. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Employment or leadership:** None declared.
**Honorarium:** None declared.
**Competing interests:** Authors state no conflict of interest.
**Ethical approval:** The local Institutional Review Board deemed the study exempt from review.

# References

1. Branford S, Cross NC, Hochhaus A, Radich J, Saglio G, Kaeda J, et al. Rationale for the recommendations for harmonizing current methodology for detecting BCR-ABL transcripts in patients with chronic myeloid leukaemia. Leukemia 2006;20:1925–30.
2. Cross NC, White HE, Müller MC, Saglio G, Hochhaus A. Standardized definitions of molecular response in chronic myeloid leukemia. Leukemia 2012;26:2172–5.
3. Branford S, Fletcher L, Cross NC, Müller MC, Hochhaus A, Kim DW, 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;112:3330–8.
4. White HE, Matejtschuk P, Rigsby P, Gabert J, Lin F, Lynn Wang Y, et al. Establishment of the first World Health Organization International Genetic Reference Panel for quantitation of BCR-ABL mRNA. Blood 2010;116:e111–7.
5. Ruiz MS, Medina M, Tapia I, Mordoh J, Cross NC, Larripa I, et al. Standardization of molecular monitoring for chronic myeloid leukemia in Latin America using locally produced secondary cellular calibrators. Leukemia 2016;30:2258–60.
6. Gabert J, Beillard E, van der Velden VH, Bi W, Grimwade D, Pallisgaard N, 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;17:2318–57.
7. Beillard E, Pallisgaard N, van der Velden VH, Bi W, Dee R, van der Schoot E, 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;17:2474–86.
8. Branford S. Monitoring and defining early response: where to draw the line? Best Pract Res Clin Haematol 2016;29:284–94.
9. White HE, Hedges J, Bendit I, Branford S, Colomer D, Hochhaus A, et al. Establishment and validation of analytical reference panels for the standardization of quantitative BCR-ABL1 measurements on the international scale. Clin Chem 2013;59:938–48.
10. Branford S. Molecular monitoring in chronic myeloid leukemia-how low can you go? Hematology Am Soc Hematol Educ Program 2016;2016:156–63.
11. Cross NC, Hochhaus A. Deep molecular response in chronic myeloid leukemia. Leukemia 2016;30:1632.
12. Cross NC, White HE, Ernst T, Welden L, Dietz C, Saglio G, et al. Development and evaluation of a secondary reference panel for BCR-ABL1 quantification on the International Scale. Leukemia 2016;30:1844–52.
13. Branford S, Yeung DT, Parker WT, Roberts ND, Purins L, Braley JA, et al. Prognosis for patients with CML and >10% BCR-ABL1 after 3 months of imatinib depends on the rate of BCR-ABL1 decline. Blood 2014;124:511–8.
14. Hanfstein B, Shlyakhto V, Lauseker M, Hehlmann R, Saussele S, Dietz C, et al. Velocity of early BCR-ABL transcript elimination as an optimized predictor of outcome in chronic myeloid leukemia (CML) patients in chronic phase on treatment with imatinib. Leukemia 2014;28:1988–92.
15. Cross NC. Standardisation of molecular monitoring for chronic myeloid leukaemia. Best Pract Res Clin Haematol 2009;22:355–65.
16. Dulucq S, Etienne G, Morisset S, Klein E, Chollet C, Robbesyn F, et al. Impact of second decline rate of BCR-ABL1 transcript on clinical outcome of chronic phase chronic myeloid leukemia patients on imatinib first-line. Ann Hematol 2019;98:1159–68.