Potency Evaluation of Recombinant Human Erythropoietin in Brazil: Assessment of Reproducibility Using a Practical Approach

Michele Cardoso do Nascimento, Clarice Lima do Canto Abreu, Rodrigo Netto Costa, Wlamir Correa de Moura and Isabella Fernandes Delgado*
Instituto Nacional de Controle de Qualidade em Saúde; Fundação Oswaldo Cruz; Rio de Janeiro - RJ - Brasil

ABSTRACT
In this study, we compared the results of potency determination of recombinant human erythropoietin (rhEPO) obtained between 2010 and 2012 by the National Institute of Quality Control in Health (INCQS/Fiocruz), i.e., the National Control Laboratory (NCL), and by a manufacturer of rhEPO. In total, 47 different batches of commercially prepared rhEPO (alpha isoform) were analyzed. All results, including those of the control and warning limits, remained within the limits recommended by European Pharmacopoeia (Ph. Eur.). All relative error (RE) values were less than ±30%, whereas most were approximately ±20%. Applying the Bland–Altman plot, only two of 47 values remained outside the limits of agreement (LA). In addition, agreement of potency determination between INCQS and the manufacturer coefficient of variation of reproducibility (% CVR) was considered satisfactory. Taken together, our results demonstrate (i.) the potency assay of rhEPO performed at INCQS, is standardized and controlled, (ii.) the comparison of our results with those of the manufacturer, revealed an adequate inter-laboratory variation, and (iii.) the critical appraisal proposed here appears to be a feasible tool to assess the reproducibility of biological activity, providing additional information regarding monitoring and production consistency to manufacturers and NCLs.

Key words: Recombinant Human Erythropoietin, Biological, Quality Control, Sanitary Surveillance

INTRODUCTION

Erythropoietin (EPO) is a glycoprotein hormone essential to life. EPO is produced in adults mainly by the renal cortex, and its main function is regulation of erythropoiesis (Jacobson et al. 1957; Fried 1972; Zanjani et al. 1981; Fisher et al. 1996; Schmidt et al. 2003). In 1977, EPO was extracted and purified from the urine of anemic patients. With the advent of recombinant DNA technology, on the basis of amino acid sequence data, the human EPO gene was cloned and used to obtain recombinant human EPO (rhEPO). This was a technological breakthrough that revolutionized the treatment of anemia (Miyake et al. 1977; Lin et al. 1985).

Before the development of rhEPO, blood transfusion was the most common treatment for anemic patients. However, since its advent, rhEPO has been widely used in clinical practice to reduce the need for blood transfusions in surgical procedures (Faris et al. 1996) and in the treatment of anemia of various etiologies (Rizzo et al. 2001; Saag et al. 2004).

A factor of major importance resulting from the use of rhEPO is the improvement in quality of life,
and in general, increased survival, as well as reduced fatigue. Its use allows normalization of the appetite of anemic patients and accelerates their return to routine activities (Aapro et al. 2006; Guan and Chen 2008).

The patent of the first commercially available EPO expired in 2007, leading to the emergence of similar versions of EPO in the global market, called biosimilars. This occurrence raised a concern among regulatory agencies, particularly with regard to the efficacy and safety of these products, because they are not identical to the original product (WHO 2010; Brinks et al. 2011).

Given the range of products now available on the market, the scope of therapeutic indications, and the characteristics of patients using rhEPO, the effective quality control of these products is of major importance prior to their market entry. Accurate quantification of potency or biological activity is a very important laboratory evaluation because it is related to the effectiveness of the final product and is recommended by the European Pharmacopoeia (Ph. Eur.) (Council of Europe 2011). As described in this compendium, this assessment can be performed by two different methods: (i) using polycythemic mice (method A) or (ii) using normocythemic mice (method B) (Council of Europe 2011).

Method B is normally the method of choice because method A inflicts greater stress and suffering on the animals that must be kept for long periods in hypobaric chambers and exposed to radioisotopes (ECVAM 2002). Therefore, different strains of normocythemic mice have been tested to evaluate rhEPO potency, such as CF1, Balb/c, Swiss Webster, NIH, C57BL6 and B6D2F1 strains (Albertengo et al. 1999; Ramos et al. 2003; Schmidt et al. 2003; Barth et al. 2008; Council of Europe 2011; Silva et al. 2013).

The process of batch release of certain biological products, such as rhEPO, is generally characterized by the use of a large number of animals. This occurs, among other reasons, because of the inherent variability of those biological assays, which often leads to the need for repetitions. This warrants, the search for new strategies, that could improve the reliability of in vivo results and consequently reduce the incidence of invalid assays. This is particularly important currently because the search for alternatives to animal testing is a priority worldwide (De Mattia et al. 2011).

In this study, we compared the results of potency determination obtained between 2010 and 2012 by the National Institute of Quality Control in Health— (INCQS/Fiocruz), i.e., the National Control Laboratory (NCL), and by a manufacturer of rhEPO. Through this critical appraisal, we aimed to refine the quality control process for this important biological agent, which is widely distributed by the Brazilian Ministry of Health.

**MATERIAL AND METHODS**

**Pharmaceutical products and reagents**

The working reference material, with a potency of 3,773 International Units (IU)/mL, known as MRT(B)rhEPO/0208, was previously established in a collaborative study conducted by INCQS/Fiocruz (Rio de Janeiro, Brazil), the Institute of Immunobiological Technology (Bio-Manguinhos/Fiocruz, Rio de Janeiro, Brazil) and the Centre of Molecular Immunology (CIM, Havana, Cuba). This reference material was provided to INCQS by Bio-Manguinhos and was used in each assay performed.

In total, 47 different batches of rhEPO (alpha isoform), all from the same producer, were identified by Arabic numerals from 1 to 47. All preparations were within their shelf-life and had a concentration of 4,000 IU/mL.

The following reagents were also used: methylene blue, sodium citrate, sodium chloride, and sodium phosphate dibasic (Merek, Darmstadt, Germany); bovine serum albumin, fraction V (Sigma-Aldrich, Missouri, USA); heparin (5,000 IU/mL) (Eurofarma, São Paulo, Brazil); lysis solution, consisting of potassium cyanide and surfactant solution (CELM, São Paulo, Brazil); and oxybuprocaine (Latinofarma, São Paulo, Brazil). The aforementioned reagents were of the highest purity available from commercial sources.

**Laboratory animals**

Female normocythaemic B6D2F1 mice from the Centre of Laboratory Animal Breeding (Fiocruz/Brazil) were housed under controlled conditions (room temperature, 21 ± 2°C; humidity, 55 ± 10%; artificial illumination, 12 h/day) with ad libitum access to water and food. For the assays, the animals were age-matched, usually at 6–8 weeks, with a body weight ranging from 15 to 17 g on the day of injection, as recommended by Ph. Eur. (Council of Europe 2011). The experimental...
protocol was followed in accordance with the EU Directive 2010/63/EU for animal experiments and was previously approved (license LW-14/11) by the Ethics Committee on Animal Use (Fiocruz).

**Biological assay**

All assays were performed from 2010 to 2012. To develop a 3 × 3 test with six points (3 dose of standard × 3 dose of samples on test), experiments were performed using six animals (randomly assigned) per dose and three doses per preparation (the test sample and reference standard). Standard and test samples were diluted to appropriate concentrations with PBS containing 0.1% bovine serum albumin. A single dose of 30, 90, or 270 IU/0.2 mL per mouse was inoculated by subcutaneous injection on day 0. On day 4, mice were bleed from the orbital venous sinus, under a topical ophthalmic anesthesia (oxybuprocaine) and blood (200 μL) was drawn from each mouse using a Pasteur pipette with 10 μL of heparin (5,000 IU/mL) for the selective hemolysis counting method. Blood was kept at 2-8 °C. The animals were euthanized by inhalation of carbon dioxide. Forty microliters of blood from each individual animal was transferred to a series of labeled tubes containing 120 μL of a mixture of two solutions, one of 0.38 g sodium citrate in 10 mL of distilled water and another of 0.12 g of methylene blue in 5 mL of sodium chloride. The mixture was incubated in a water bath at 37°C for 1 h. Then, 40 μL of hemolyzing solution comprising of 120 μL of lysis solution plus 2,580 μL of sodium chloride were added and left at room temperature for two cycles of incubation for 3 min after homogenization. Then, 10 μL of the hemolyzed mixture were transferred to another series of assay tubes containing 1,960 μL of sodium chloride solution. After homogenization, 10 μL samples of the suspensions was transferred to a hemocytometer, and the reticulocytes were counted under a microscope (400 × magnification) and expressed as an absolute value.

**Manufacturer results**

The results from the producing laboratory, for the same 47 batches analyzed by INCQS, were obtained after analysis of the summarized production and control protocol. The manufacturer performed each potency determination as recommended by Ph. Eur. (Council of Europe 2011), and as well as INCQS, it used B6D2F1 mice to assess rhEPO potency. All assays were performed from 2010 to 2012.

**Statistical analysis**

Statistical analysis of the data was performed by the parallel line method as described in Ph. Eur. (Council of Europe 2012) using CombiStats™ from European Directorate for the Quality of Medicines & HealthCare - EDQM (EDQM 2013). The validity of each assay was demonstrated by analysis of variance (ANOVA), evaluating the significance of linear regression and deviations from linearity and parallelism, such that the test was considered valid when the regression was significant ($P < 0.05$) and there were no statistically significant deviations from linearity and parallelism ($P > 0.05$). The biological potency and confidence limits of the samples, which must lie between 80% and 125% and between 64% and 156%, respectively, were expressed in relation to the potency declared by the producer. The Kolmogorov–Smirnov (KS) test was applied to check if the data fit the normal distribution.

**Control charts**

The principle of monitoring, which is a very important step in the program of quality assurance, tests how much data can be considered to have come from a population and indicates whether the processes of production and testing are consistent. Monitoring can be accomplished using control charts (Hendriksen et al. 2008), and its application to the quality control of vaccines was described 60 years ago by Batson and colleagues (Batson et al. 1951). Trend analysis of the results obtained by INCQS, including the behavior of the reference standard, as well as comparisons between their results and those obtained by the producer, should be performed (Hendriksen et al. 2008; WHO 2010). Using SPC Explorer RT™ (Quality America Inc.; EUA), the control charts were plotted for both INCQS and the manufacturer results of the 47 samples tested. The warning ($±2$ SD), control limits ($±3$ SD) and mean using the first 20 results was calculated (Bastson et al. 1951; Hendriksen et al. 2008; Pyzdek 2008; WHO 2013).

**Relative error**

The relative error (RE) was calculated to evaluate the agreement between the two laboratories. RE indicated the bias, which represents the difference.
between the expected result of a test or measurement and a true value (ISO 2006). To calculate the bias between the results obtained by the two laboratories, the potency values obtained by INCQS were considered the true value and the difference between these results and those of the producer for each pair of results was considered RE. WHO recommends combining results when multiple assays are performed. Once the range of accepted results is from 80 to 125% of the declared potencies (Council of Europe 2011), an acceptable variation of 45% is officially created for trueness (bias) and precision. Considering this wide variation an empirical acceptance criterion of 30% for trueness and precision was adopted.

Bland–Altman plot
According to Bland and Altman (1986), measuring the differences between two methods for each sample and testing them against the mean is the best way to evaluate the relationship between the methods. The dispersion of the differences along the zero line is an indication of the agreement between the two methods (Bland and Altman 1986). Thus, the Bland–Altman plot was applied to evaluate the values obtained by INCQS and the manufacturer. It was created using GraphPad Prism™ by calculating the average, the difference between each pair of data points, the bias, and the limits of agreement (LA). The differences plotted against the mean were expected to be within the LA calculated.

Homogeneity of results
To evaluate the homogeneity of results obtained by INCQS and the manufacturer, the chi-square test was applied to calculate the difference between the groups. This was assessed using CombiStats™ (EDQM 2013), as described in Ph. Eur. (Council of Europe 2012) and United States Pharmacopoeia (USP) (USP 2012a).

Combination of assays results
WHO recognizes that when multiple assays are performed, the results must be combined. This is also described by both Ph. Eur. and USP (Council of Europe 2012; USP 2012a).

Reproducibility
Because a conventional collaborative study was not conducted, in which the same samples would be analyzed in matrices of tests in different laboratories, a practical approach was adopted to determine the reproducibility (or precision between laboratories). The potency determination results of 47 batches from the same manufacturer and those obtained by INCQS using the B6D2F1 strain were compared by establishing the variation between the results of the two laboratories for the different samples tested. This was accomplished by averaging the variances between the results obtained by INCQS and the manufacturer to determine the coefficient of variation of reproducibility (%CV_R). It was adopted the acceptance criterion of 30% as the maximum acceptable variation between the results of the two laboratories.

RESULTS
The potency determination values are presented in control charts in Figure 1. All the results remained within the limits recommended by Ph. Eur. (Council of Europe 2011), i.e., 80%–125% of the labeled potency and of the control and warning limits calculated.
In all groups of data, normal fit was demonstrated by the KS test ($P > 0.05$). Thus, no transformation was required. All data are obtained from single assay results (assay repetition was not required). The dispersion of RE of the potency determination values related to the 47 batches analyzed in both laboratories is shown in Figure 2. All RE values were less than ± 30%, whereas most ranged between ± 20%.

**Figure 2** - Relative Error between the potency values determined by manufacturer and INCQS.

Bland–Altman plot differences between the averages determined by INCQS and the manufacturer for each sample tested is shown in Figure 3. Bias was 1.851 and LA was from -20.04 to 23.74. Only two values remained outside LA.

**Figure 3** - Bland–Altman plot showing the difference between the potency values determined by the manufacturer and INCQS. LA: Limits of Agreement.

With CombiStats™ (EDQM 2013), it was possible to evaluate the homogeneity of these results using the chi-square test between each pair of potency determination values and confidence limits obtained by INCQS and the manufacturer. Only one result was considered heterogeneous ($P < 0.05$), whereas the other results were considered homogeneous ($P > 0.05$), i.e., 97.87% of the results were homogeneous.

**DISCUSSION**

Since its discovery, purification, cloning, and production by recombinant DNA technology, giving rise to rhEPO, EPO has been widely studied (Miyake et al. 1977; Lin et al. 1985; Costa et al. 2010; Brinks et al. 2011). When selecting assays to be used for the potency evaluation of rhEPO, the complexity of biological medicine, with a heterogeneous glycosylation pattern, must be considered because the sialic acid content can significantly influence its metabolism and biological activity. Accordingly, and given this complexity, there are still no *in vitro* models for the evaluation of biological activity of rhEPO, and internationally, it is recommended to use *in vivo* assays to determine its potency (Albertengo et al. 1999; Council of Europe 2011).

All assays performed by INCQS and all results from the manufacturer complied with the validity criteria (regression, linearity, and parallelism) and were within the limits established by Ph. Eur. (Council of Europe 2011), i.e., potency between 80% and 125% limits and between 64% and 156% of rated potency. In all groups of evaluated data, the KS test provided no statistical evidence of a lack of fit to the normal distribution ($P > 0.05$). Although it is possible to combine valid results obtained in independent potency determination tests performed upon batch release of rhEPO (Council of Europe 2012; USP 2012a; WHO 2013), in this study, all the results of INCQS were obtained from single assays because there was no need for combination trials of the satisfactory results in the samples tested. These data differ from those of previous studies, where it was necessary to combine two or even three independent experiments to obtain satisfactory results for the same sample in different strains, such as in Swiss Webster, B6D2F1, CF1, or BAlb/c mice (Albertengo et al. 1999; Ramos et al. 2003; Schmidt et al. 2003; Lopes 2004; Barth et al. 2016).
This is probably due to the optimization achieved at our institute in terms of standardization and control factors (such as the range of body weight and age of the animals, ambient humidity and temperature of the animal facilities, precision in sample preparation and reading of reticulocytes, and use of calibrated instruments) that could affect the outcome of a trial, thereby reducing the variability inherent in the assay.

The variability of *in vivo* assays is well known and is much greater than that expected for other types of testing, such as chemical assays. According to the WHO (1997), a bioassay—whether *in vitro* or *in vivo*—may have a variability of more than 50%. This has led to the adoption of the relative potency methodology because an absolute measurement is more variable than the measurement of relative activity in relation to a certain standard (USP 2012b).

Another approach, which was also used in the present study, concerns the assessment of the applicability of control charts as an additional tool for quality control and assesses the consistency of biological assays results. Control charts for individual values are statistical tools used to assess the central tendency of a process over time and to support preventive and/or corrective actions when necessary (Hendriksen et al. 2008).

The control charts (Fig. 1) showed that the assays remained consistent over time, in both INCQS and the manufacturer. Thus, nonrandom trends were not observed. This means that the test process is under statistical control, and the accuracy and precision of the results are maintained, producing results in which the mean and standard deviation are predictable when applied repeatedly to the same material over time. In this situation, the process of quality control is subjected only to random errors from multiple small common causes inherent in the test system.

Comparison of our test results with those of the manufacturer revealed that RE was within ±30%. This variation is expected and inherent in biological assays. Moreover, ANOVA of the data produced homogeneous results, and no significant difference was observed between groups ($P > 0.05$).

After applying the Bland–Altman plots to these same data, only two values remained outside LA, demonstrating that the results were consistent between the two laboratories, as shown in Figure 3.

Assessment of the homogeneity of these results using the chi-square values between each pair of potency and confidence limits revealed that only one pair of results was considered heterogeneous ($P < 0.05$), whereas all other results were considered homogeneous ($P > 0.05$).

A practical approach was developed to evaluate the reproducibility of the assay (precision between laboratories) because a conventional collaborative study was not conducted. This was done by determining $\% CV_R$, which considers the average of the variances.

When comparing the results of potency obtained by INCQS with those obtained by the producer, $\% CV_R$ was 8.97%, which was below the acceptance criteria adopted (i.e., 30%). However, the acceptance range of the assay would allow testing differences up to 45% because the acceptable range for potency values is 80%–125% and the assays were performed in different laboratories.

In summary, our results demonstrate that the process of quality control of rhEPO in Brazil, performed by INCQS, in relation to its biological activity, is standardized and controlled. This ensures the quality of results in each batch released. In addition, the critical approach proposed here seems to be a feasible tool to assess reproducibility of biological activity, providing additional information regarding monitoring and production consistency to manufacturers and NCLs.

ACKNOWLEDGEMENTS

This work is part of the dissertation of Michele Cardoso do Nascimento presented in the Master’s Course of Sanitary Surveillance of INCQS/Fiocruz. Isabella Fernandes Delgado received a research study grant from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (DT-II/CNPq).

REFERENCES


Received: July 02, 2014; Accepted: December 02, 2014.