

## Research Article

# Incurred Sample Reanalysis: Time to Change the Sample Size Calculation?

Piotr J. Rudzki,<sup>1,3</sup>  Przemysław Biecek,<sup>2</sup>  and Michał Kaza<sup>1</sup> 

Received 23 November 2018; accepted 28 December 2018; published online 11 February 2019

**Abstract.** Reliable results of pharmacokinetic and toxicokinetic studies are vital for correct decision making during drug discovery and development. Thus, ensuring high quality of bioanalytical methods is of critical importance. Incurred sample reanalysis (ISR)—one of the tools used to validate a method—is included in the bioanalytical regulatory recommendations. The methodology of this test is well established, but the estimation of the sample size is still commented on and contested. We have applied the hypergeometric distribution to evaluate ISR test passing rates in different clinical study sizes. We have tested both fixed rates of the clinical samples—as currently recommended by FDA and EMA—and a fixed number of ISRs. Our study revealed that the passing rate using the current sample size calculation is related to the clinical study size. However, the passing rate is much less dependent on the clinical study size when a fixed number of ISRs is used. Thus, we suggest using a fixed number of ISRs, e.g., 30 samples, for all studies. We found the hypergeometric distribution to be an adequate model for the assessment of similarities in original and repeated data. This model may be further used to optimize the sample size needed for the ISR test as well as to bridge data from different methods. This paper provides a basis to re-consider current ISR recommendations and implement a more statistically rationalized and risk-controlled approach.

**KEY WORDS:** incurred sample reanalysis (ISR); bioanalysis; hypergeometric distribution; bioanalytical method validation; bridging data.

## INTRODUCTION

Reliable results of pharmacokinetic and toxicokinetic studies are vital for correct decision making during drug discovery and development. Thus, assuring high quality of bioanalytical methods is of critical importance. The American Association of Pharmaceutical Scientists (AAPS) and the US Food and Drug Administration (FDA) were the driving forces behind discussions on the bioanalytical method validation in the 1990s (1). Both AAPS and FDA are constantly involved in the evolution of bioanalytical requirements, including incurred samples reanalysis (ISR) (2,3). Professional organizations like the European Bioanalysis Forum have also presented their opinions on the test (4). Finally, the ISR was included in the bioanalytical regulatory recommendations by the European Medicines Agency (EMA) (5), the Health Canada (6), and the FDA (7). Although the ISR is now part of the regulatory documents, the topic is still under

much discussion in the bioanalytical and pharmaceutical community (8–16).

One of the most debated aspects of the ISR test is the estimation of its sample size. Originally, Rocci *et al.* proposed a fixed number of ca. 20 samples, which was argued to detect a 20% difference for small molecules with 0.8 power and 0.05 type I error (17). The European Bioanalysis Forum agreed with Rocci *et al.* and suggested a fixed number of 20–50 samples per study (4), but did not justify it statistically. Enhanced statistical considerations and the simulations presented by Hoffman revealed that for accurate (bias = 0%) and precise (CV = 10%) methods, 40 samples allowed correct decision making (18). Thway *et al.* reviewed their studies on macromolecules and used simulations to investigate the influence of the method precision on the probability of passing the ISR test (19). They revealed that 40 samples are enough to pass a 30% difference criterion with a probability close to 1, whereas the precision did not exceed 15%. The Global Bioanalysis Consortium suggested, in turn, to reanalyze the fixed ratio of 5% of clinical samples that is equal to the ratio of quality control (QC) samples in a bioanalytical batch (9). More recently, Subraminiam *et al.* presented a comprehensive ISR simulation for small molecules using various combinations of precision and bias (10). It revealed that if 20 ISRs are analyzed and the method's precision up to 15% is combined with the bias not exceeding 10%, then over

<sup>1</sup> Pharmacokinetics Department, Pharmaceutical Research Institute, 8 Rydygiera Street, 01-793, Warsaw, Poland.

<sup>2</sup> Faculty of Mathematics and Information Science, Warsaw University of Technology, 75 Koszykowa Street, 00-662, Warsaw, Poland.

<sup>3</sup> To whom correspondence should be addressed. (e-mail: p.rudzki@ifarm.eu)

0.85 of the studies pass the ISR test. All these simulations contributed to the knowledge on how different factors influence the ISR sample size. They also suggest that for reproducible methods, only 40–50 samples are enough to meet the ISR passing rate over 99%. Yet, some of the assumptions seem questionable, like no bias (18,19) or a narrow concentration range (10,19). Moreover, we have recently suggested that the number of ISRs recommended by FDA and EMA (5,7) does not seem to be well matched with the acceptance criteria (12).

To illustrate the assumptions-free approach, let us gamble for a moment. Imagine an urn with green and red balls. Our goal is to guess if the ratio of green balls to the red ones is not lower than 2:1. How many balls do we need to pick to be confident that the ratio would be 2:1? Does our confidence depend on the number of balls in the urn or rather on the true ratio of the balls? Now, let us replace ball picking with a bioanalytical method, and green balls with the passed ISR pair. Then, we can rephrase the question: how many ISRs should we analyze to confirm the method's reliability? Does it depend on the clinical study size or rather on the method's performance itself? Based on the urn example, the hypergeometric distribution may help us answer these questions. This theoretical distribution is used in physics, biology, medicine, and chemistry, for example to study signal-to-noise ratios (20), as well as in the mass spectrometric

identification of proteins (21,22), phospholipids (23), and elemental composition of unknown compounds (24). This is the first study that uses the hypergeometric distribution to evaluate the ISR sample size.

In this paper, we have aimed to provide researchers and regulators with a model for estimating and optimizing the number of ISRs needed to prove the reliability of a bioanalytical method. We have applied the hypergeometric distribution to evaluate ISR test passing rates in different clinical study sizes.

## MATERIALS AND METHODS

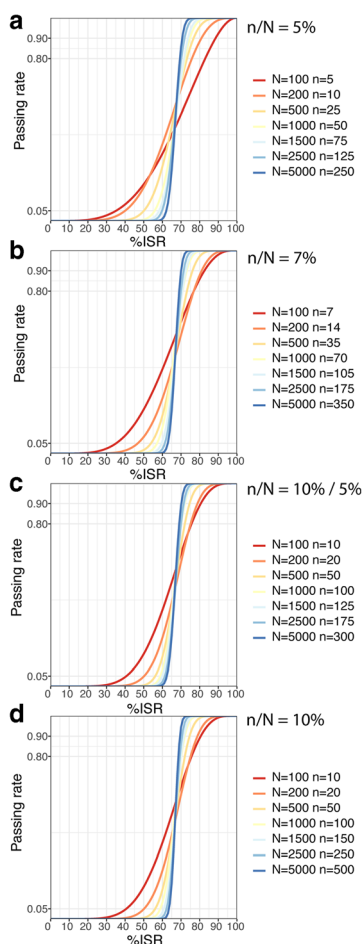
The symbols and terms used in this manuscript are defined in Table I. Calculations and Figs. 1, 2, 5, and 6 have been generated in the R version 3.5.0 (25). For Figs. 1 and 2, we have selected a red-yellow-blue color scheme for seven diverging percentage difference classes in a colorblind-safe mode (26). Figures 3 and 4 have been generated in Microsoft Office Excel 2007.

We have assumed that the population size ( $N$ ) is typically between 100 and 5000 clinical samples, but in some calculations we have also used smaller sample sizes of 20 and 50. We have tested fixed  $n/N$  ratios as recommended at 5% (9), at 7% (27), and at 10% for comparison. We have also tested the reference two-step fixed  $n/N$  ratio recommended by the regulatory authorities: 10% up to 1000 samples and then 5% (5,7). A fixed

**Table I.** Symbols and Terms Used

Symbol or term	Hypergeometric distribution	ISR test	Values tested and/or calculation method <sup>a</sup>
$N$	Size of the population	Study sample size—number of unique biological samples in a clinical study	(20), (50), 100, (200), (250), 500, 1000, 1500, 2500, and 5000
$n$	Number of the experiments	Number of ISRs	Fixed number: 10, 20, (30), 50, and 100 or fixed ratio: 5% $\cdot N$ (9), 7% $\cdot N$ (27), 10% $\cdot N$ , 10% $\cdot N$ for the studies with $N \leq 1000$ and 10% + 5% $\cdot (N-1000)$ for the studies with $N > 1000$ (5, 7)
$K$	Number of successes in the population	Number of ISR pairs meeting %difference criteria if all samples from the clinical study have been analyzed	$K = p \cdot N$ $K \in [0, 1, \dots, N]$
$k$	Number of successes in $n$ experiments	Number of ISR pairs meeting %difference criteria observed in the reanalyzed samples	$k \in [0, 1, \dots, n]$
$p = \%ISR$	Success rate	true percentage of ISR pairs meeting %difference criteria (when all samples have been reanalyzed)	$p \in [0; 100\%]$ $p = \frac{K}{N} \cdot 100\%$
$\hat{p} = \%isr$	Estimated success rate	The estimated percentage of ISR pairs meeting %difference criteria (when a portion of the samples has been reanalyzed)	$\hat{p} \in [0; 100\%]$ $\hat{p} = \frac{k}{n} \cdot 100\%$
Passing rate	–	Probability of passing the ISR test	Calculated using the hypergeometric distribution passing rate $\in [0; 1]$
%difference	–	Percentage difference between the original concentration and the concentration measured during the repeat analysis	%difference = $\frac{\text{repeat} - \text{original}}{\text{mean}} \cdot 100\%$ (7)

<sup>a</sup> Values in brackets were used in selected tests only



**Fig. 1.** Cumulative distribution function for different study sample sizes ( $N$ ), when the ratio of the number of ISRs to study the sample size ( $n/N$ ) is fixed at 5% (a), 7% (b), 10% up to 1000 samples, and then 5% (c) and 10% (d)

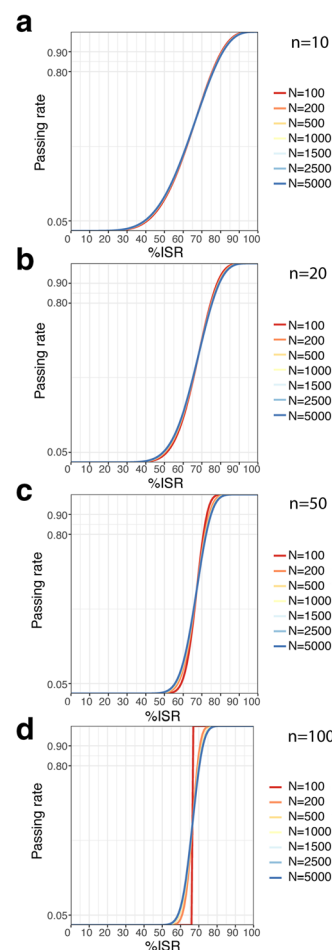
number of 20 and 50 ISRs ( $n$ ) is based on the previously reported assumptions (4,17). We have used the term “ISR pair” with reference to the related original and repeat values.

We have classified the bioanalytical methods as:

- non-reproducible (when the passing rate is not higher than 0.05),
- reproducible (when the passing rate is not less than 0.80),
- and quasi-reproducible (when the method does not meet non-reproducible nor reproducible criteria).

Note that the classification of the methods with the same %ISR—i.e., true percentage of the ISR pairs meeting %difference criteria when all samples are reanalyzed—may vary depending on  $N$  and  $n$ .

We have assumed that the test outcome for each ISR pair is dichotomous, i.e., belonging to one of two mutually exclusive categories: a success (if an ISR pair meets the %difference criteria) or a failure (if an ISR pair does not meet the %difference criteria). The sampling from a finite population has been carried out without replacement, that is why the hypergeometric distribution is more appropriate than the binomial distribution. The type I error ( $\alpha$ ) of 0.05 and power



**Fig. 2.** Cumulative distribution function for different study sample sizes ( $N$ ), when the number of ISRs ( $n$ ) is fixed at 10 (a), 20 (b), 50 (c), and 100 (d)

of 80% have been used as suggested by Rocci et al. (17). Additionally, we have used the power of 90%.

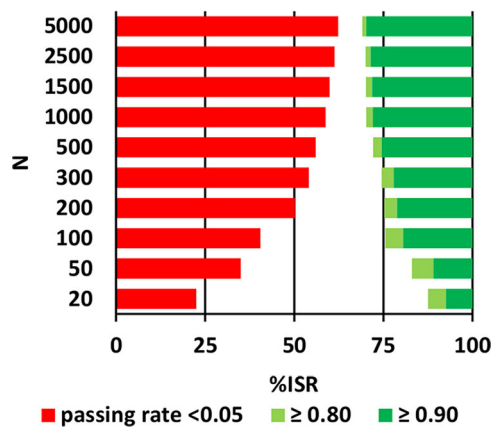
The acceptance criteria were in line with the FDA (7) and the EMA (5) recommendations on the bioanalytical method validation. The %difference (Table I) should be within  $\pm 20\%$  of the mean concentration for small molecules and within  $\pm 30\%$  for large molecules. The percentage of the ISR pairs meeting the %difference criteria (%*isr*) should be at least 67%.

## RESULTS

### Fixed Ratio of the Number of ISRs to Study Sample Size ( $n/N$ )

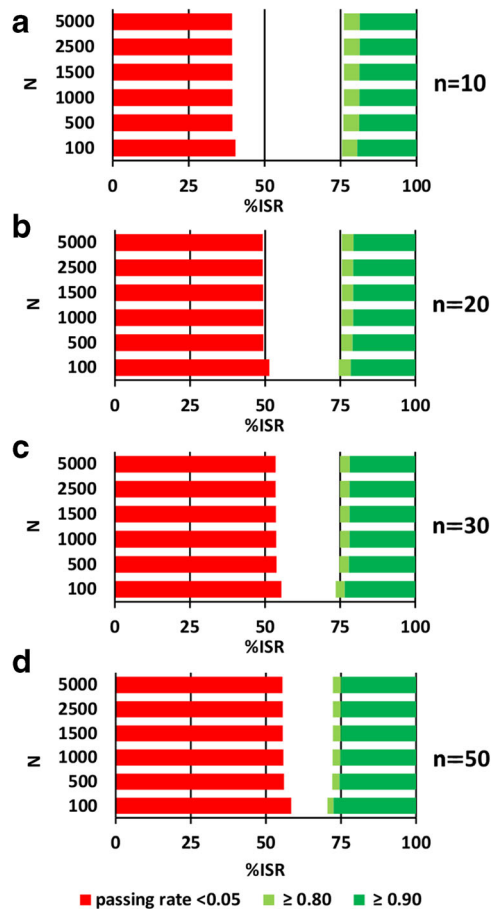
We have started from the scenario when  $n/N$  is fixed. This is in line with the current regulatory practice which uses a two-step fixed ratio depending on the clinical study size:  $n/N = 10\%$  up to 1000 clinical samples and then  $n/N = 5\%$  (5,7). We have also evaluated one-step fixed  $n/N$  at 5% (9), 7% (27), and 10%. For each  $n/N$ , we have plotted the passing rate vs. %ISR (Fig. 1) for selected  $N$  values.

Generally, all plots look similar regardless of  $n/N$  used (Fig. 1). Except for  $N$  of 100 and 200, the passing rates are comparable for all  $n/N$  ratios, including the passing rate (I) below



**Fig. 3.** Differentiation of non-reproducible (red) and reproducible (green) methods for different study sample sizes ( $N$ ) using the current regulatory ISR sample size (5,7)

0.05 when  $\%ISR \leq 50$ , (II) over 0.80 when  $\%ISR \geq 75$ , and (III) over 0.90 when  $\%ISR \geq 80$ . For  $N \leq 200$ —or for  $n$  ranging from 5 to 20—cumulative distribution functions look flatter. The results suggest that in this scenario non-reproducible and reproducible methods are better discriminated as the sample size increases.



**Fig. 4.** Differentiation of non-reproducible (red) and reproducible (green) methods for different study sample sizes ( $N$ ), when the number of ISRs ( $n$ ) is fixed at 10 (a), 20 (b), 30 (c), and 50 (d)

**Fixed Number of ISRs ( $n$ )**

Another approach to the ISR sample size is a fixed number of ISR pairs apart from the study size (4,17). So, we have fixed  $n$  at the following values: 10, 20, 50, and 100. We have combined each  $n$  with the selected  $N$  values and then we have plotted the passing rate vs.  $\%ISR$  (Fig. 2).

Surprisingly and contrary to the fixed  $n/N$  (Fig. 1), nearly all curves for particular  $n$  are overlapping (Fig. 2). Only when both  $n > 20$  and  $N = 100$  the cumulative distribution functions look steeper (Fig. 2c, d,  $N = 100$ ). Apart from these two exceptions—both far exceeding the currently recommended  $n/N$  of 10% (5,7)—the passing rate seems independent of the study sample size.

A more detailed look at the plots shows that when  $n$  increases, the cumulative distribution functions are steeper in the  $\%ISR$  range of 50–70%. Thus, as could be expected, the reproducible methods are better distinguished from the non-reproducible ones when  $n$  increases. But to what extent should they be distinguished?

**To What Extent Should the Reproducible Methods Be Distinguished from the Non-reproducible Ones?**

We needed two steps to answer this question. Firstly, we selected critical passing rates of  $\leq 0.05$  for non-reproducible methods (equal to the assumed acceptable type I error) as well as the passing rates of  $\geq 0.80$  and  $\geq 0.90$  for reproducible methods (equal to the typical values of power in bioequivalence studies (28)). As a reference, we calculated  $\%ISR$  for each of the above passing rates (Table II) using the current regulatory ISR sample size ( $n$ ) (5,7). Data presented in Table II and in Fig. 3 both confirm that using a two-step fixed ratio leads to different  $\%ISR$  needed to get the same passing rate for different clinical study sizes ( $N$ ). For example, in order to achieve the passing rate not exceeding 0.05 when  $N = 20$ , the method may have very low reproducibility ( $\%ISR = 22.5\%$ ). But when  $N = 5000$ , then reproducibility may be quite similar to the acceptance criteria ( $\%ISR = 62.3\%$ ). Thus, the current ISR sample size ( $n$ ) leads to the acceptance criteria dependent on the clinical study size ( $N$ ).

In the second step, we calculated  $\%ISR$  for each of the critical passing rates using fixed  $n$  (Fig. 4, Table III). Contrary to the fixed  $n/N$  ratio, we observed a similar  $\%ISR$  needed to get the same passing rate for different clinical study sizes ( $N$ ). For example, when  $n = 30$  is used to achieve the passing rate not lower than 0.80, for  $N = 100$  the method should have the reproducibility of  $\%ISR = 73.5\%$ , while for  $N = 5000$  the reproducibility of 1.4% higher ( $\%ISR = 74.9\%$ ) is necessary. Thus, the fixed  $n$  leads to the acceptance criteria much less dependent on the clinical study size ( $N$ ).

As expected, the higher the  $n$ , the better reproducible methods are distinguished from the non-reproducible ones. But to what extent should they be distinguished? To figure out the answer, we added a pair of assumptions. Firstly, when 1 of every 2 samples meets the ISR acceptance criteria ( $\%ISR = 50\%$ ), then the method is non-reproducible. Secondly, when 3 of every 4 samples meet the ISR acceptance criteria ( $\%ISR = 75\%$ ), then the method is reproducible. For these assumptions, the hypergeometric distribution suggests that  $n = 30$  is the right selection. For such  $n$ , the  $\%ISR$  over 50% is needed to achieve the passing rate of 0.05. For the sample sizes of up to 5000, the  $\%ISR$  below 75% is enough to get the passing rate of 0.80 (Fig. 5, Fig. 7 in Appendix I).

**Table II.** Theoretical %ISR Calculated for Different Passing Rates Using the Current Regulatory ISR Sample Size (5,7)

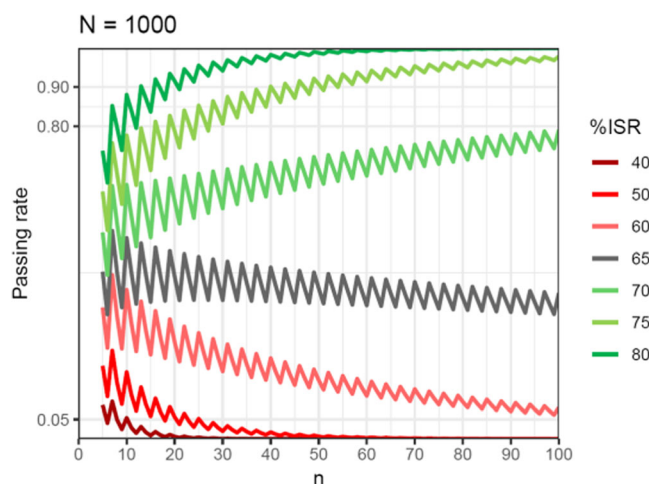
<i>N</i>	<i>n</i>	Passing rate ≤ 0.05	Passing rate ≥ 0.80	Passing rate ≥ 0.90
20	2	22.5%	87.5%	92.5%
50	5	35.0%	83.0%	89.0%
100	10	40.5%	75.5%	80.5%
200	20	50.2%	75.3%	78.8%
300	30 <sup>a</sup>	54.1%	74.5%	77.9%
500	50	56.0%	72.1%	74.5%
1000	100	58.8%	70.2%	72.0%
1500	125	59.9%	70.1%	71.8%
2500	175	61.3%	70.0%	71.4%
5000	300	62.3%	69.1%	70.2%

<sup>a</sup> Values in italics: the lowest *n* among the selected values to meet the following criteria: %ISR > 50% needed for the passing rate of 0.05 and %ISR < 75% enough for the passing rate of 0.80

**Table III.** Theoretical %ISR Calculated for Different Passing Rates Using a Fixed Number of ISRs (*n*)

<i>N</i>	<i>n</i>	%ISR (%)		
		Passing rate ≤ 0.05	Passing rate ≥ 0.80	Passing rate ≥ 0.90
100	10	40.4	75.5	80.5
500	10	39.4	75.9	81.1
1000	10	39.4	76.1	81.2
1500	10	39.4	76.1	81.2
2500	10	39.3	76.1	81.3
5000	10	39.3	76.1	81.3
100	20	51.4	74.5	78.5
500	20	49.4	75.3	79.1
1000	20	49.4	75.4	79.3
1500	20	49.3	75.5	79.3
2500	20	49.2	75.5	79.3
5000	20	49.2	75.5	79.4
100	30 <sup>a</sup>	55.4	73.5	76.5
500	30	53.8	74.7	77.9
1000	30	53.7	74.9	78.1
1500	30	53.6	74.9	78.1
2500	30	53.5	74.9	78.1
5000	30	53.5	74.9	78.2
100	50	58.4	70.5	72.5
500	50	56.0	72.1	74.5
1000	50	55.8	72.2	74.8
1500	50	55.6	72.3	74.9
2500	50	55.6	72.3	74.9
5000	50	55.5	72.3	75.0
100	100	66.4	66.5	66.5
500	100	59.2	69.9	71.7
1000	100	58.8	70.2	72.0
1500	100	58.6	70.3	72.1
2500	100	58.6	70.3	72.2
5000	100	58.5	70.3	72.3

<sup>a</sup> The lowest *n* among the selected values to meet the following criteria: %ISR > 50% needed for the passing rate of 0.05 and %ISR < 75% enough for the passing rate of 0.80



**Fig. 5.** Passing rates for different %ISRs as a function of the number of ISRs (*n*) for the study sample size *N* = 1000

**How Does the Passing Rate Depend on %ISR?**

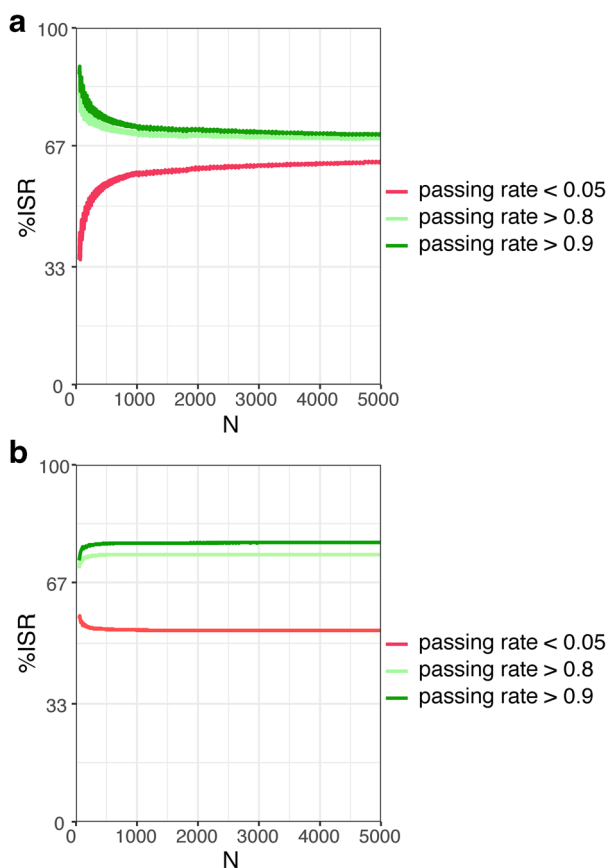
To answer this question, we compared the fixed *n/N* ratio and fixed *n* concepts using %ISR needed to achieve particular passing rates. Figure 6 confirmed previous observations for different clinical study sizes. The %ISR depends on the sample size (*N*) for the fixed *n/N*, but is much less dependent on the sample size (*N*) for the fixed *n*.

**DISCUSSION**

The goal of the ISR test is to confirm that a bioanalytical method is reliable (5,7). Thus, the probability of meeting the acceptance criteria should depend mainly—or even solely—on the bioanalytical method performance. Our new approach shows that this is not the case when the sample size is based on a fixed *n/N* ratio. The hypergeometric distribution revealed that the passing rate for a particular method performance (%ISR) is related to *N*. Surprisingly, this dependence is hard to observe for a fixed *n*.

Following the assumptions presented above, a fixed *n* = 30 seems to be the statistically right solution. The advantages of this approach over the current practice (5,7) are as follows:

- (I). passing rates are much less dependent on *N* (Fig. 6); thus, the same performance means the same probability of passing the ISR test regardless of the clinical study size,
- (II). it is simple and does not need any calculations in order to assess the ISR sample size,
- (III). non-reproducible methods are better distinguished from the reproducible ones for *N* < 300,
- (IV). fewer samples are analyzed for *N* > 300, which allows cost-effective and environmentally friendly bioanalysis in medium and large studies.



**Fig. 6.** The %ISR needed to achieve a particular passing rate for different clinical study sizes using **a** fixed  $n/N$  ratio according to the current regulatory recommendations (5,7) and **b** a fixed  $n=30$ . This figure confirms that the %ISR depends on the sample size ( $N$ ) for the fixed  $n/N$ , but is nearly independent on the sample size ( $N$ ) for the fixed  $n$

But, is the ISR sample size limited to 30 samples enough to detect problems with the method? A solely statistical approach may not be adequate to answer this question. An ISR test requires an appropriate experimental design (17,18). Samples for the test should include all phases of the study, different analytical batches, different subjects and samples stored for longer periods of time, and high and low analyte concentrations. An adequate representation of all variability factors is needed to detect problems with inhomogeneity of the sample as well as metabolism and stability issues. Thus, the proposed 30 samples is just an example, not a final solution. One should also note that the ISR test is not performed in vacuum. It is complemented by a system suitability test (SST) which confirms instrumental performance. Then, for each bioanalytical batch, the actual method performance is monitored by the calibration curve and QC samples. The suggested  $n=30$  is larger than (I) 20 ISRs initially proposed by Rocci *et al.* (17) and (II) a minimum of 20 samples proposed recently by FDA for bridging the data from multiple bioanalytical technologies (7). It is large

enough to expect valid results of statistical analysis based on the normal distribution. It is also more comparable to the sample sizes of other validation tests. One may draw a comparison to the accuracy and precision evaluation: 5 samples of high (near maximum concentration) and low (in the elimination phase) concentration, each studied in 3 separate analytical runs, gives exactly 30 samples. Should the regulatory authorities find the simple approach of a fixed  $n=30$  unsuitable, it may be somewhat extended. One possibility is an adaptive method similar to our previous concept (12).

The calculations using the hypergeometric distribution are complementary to the published simulations (10,18,19), but they have significant advantages. The novel model uses %ISR to include both random (imprecision) and systematic (bias, metabolite conversion) errors. So, the inference for many combinations of precision and bias values is avoided, which greatly simplifies the problem. We have also managed to avoid making unnecessary assumptions—like the concentration range or the distribution of the concentration for individual results (10,18,19). Thus, our calculations presented here are better suited to the real datasets than the simulations. Another novelty is defining the results of each ISR pair as dichotomous (success or failure), what makes the calculations independent of the %difference acceptance criteria. Therefore, they are valid for both small and large molecules. The hypergeometric distribution is even more universal model as it may be also used to bridge data from multiple bioanalytical technologies.

One of the limitations of this paper may lie in the assumption that %ISR is a constant. One may argue that the true value of %ISR is unknown and a particular method may have somewhat different %ISR in different studies. Due to many factors contributing to the variability, the %ISR for a particular method in a single study may vary in time. Yet, by definition, %ISR is an estimation of %ISR and we have limited the constant %ISR evaluation to a particular study. One may also suggest that a smaller number of ISRs will decrease a chance of locating problems, especially in larger studies. Each reanalysis increases the chance of unmatched results and figuring out their cause, but these opportunities are not always necessary to validate bioanalytical data (12). The hypergeometric distribution shows that an increasing number of samples over certain limit does not lead to better distinguishing of reproducible and non-reproducible methods (Fig. 6).

Bioanalysis is an important part of the drug development process. Finding an optimal ISR sample size needs appropriate balance between test ability to identify method-related problems and avoiding unnecessary analyses. The latter generate extra costs and delay research. So, creating a new performance standard for ISR may be one of the steps helping to provide patients with more timely and affordable access to new therapies, as suggested by the FDA (29). The reduction of the regulatory burden is especially anticipated for large studies, where hundreds of ISRs are being analyzed now, but even for standard-sized

studies, savings may be considerable. This paper provides a basis to re-consider the current ISR sample size calculation based on the clinical study size ( $N$ ). The proposed fixed  $n$  concept is not intended for instant practical application in the regulated bioanalysis, because it is challenging the current regulatory recommendations (5–7). However, the hypergeometric distribution has proved to be an appropriate model to help understand statistical relations between the accuracy and precision of a bioanalytical method vs. ISR test passing rates. So, the investigators interested in efficient way of validating repeatability of their non-regulated bioanalytical methods may use our approach instantly. We hope that this model may help regulatory bodies to implement more statistically rationalized and risk-controlled ISR methodology. It may be the right time for change, as the ICH is currently developing its global bioanalytical method validation guideline (30). The acceptance of the idea might need more detailed comparison of all the models used for ISR evaluation, thus future studies on this topic are invited.

## CONCLUSION

The hypergeometric distribution is an appropriate model to understand and optimize ISR sample size better. Our study revealed that the passing rates are currently related to clinical study size. Interestingly, the passing rates are much less dependent on clinical study size when a fixed number of ISR samples are used; therefore, we propose to use a constant number of samples, e.g., 30, for ISR for all studies. This paper provides a basis to re-consider ISR methodology and implement a more statistically rationalized and risk-controlled approach.

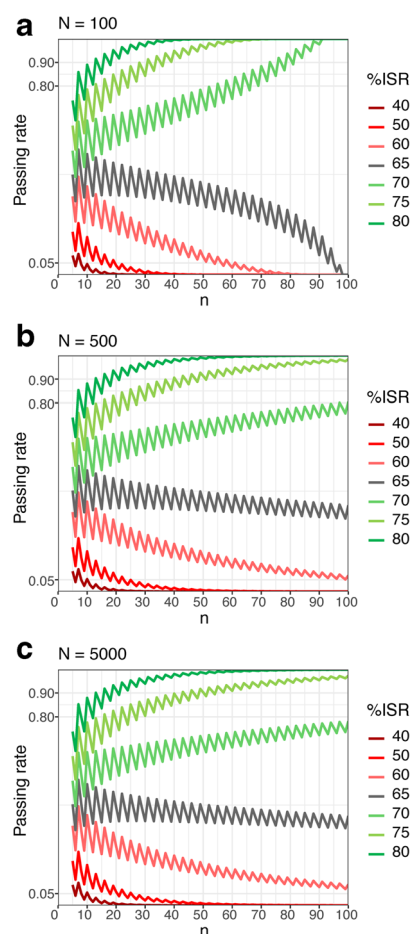
## ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. K. Buś-Kwaśnik and Ms. Edyta Gilant from the Pharmaceutical Research Institute (Poland) and Mr. Davit Sargsyan from Rutgers, the State University of New Jersey (US), for the critical reviewing of the manuscript.

## FUNDING

P. Biecek is financed by the Opus grant 2016/21/B/ST6/02176 of the National Science Center, Poland. P. Rudzki and M. Kaza are involved in the ORBIS project that received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no 778051. This Project has received supplementary funding from the Ministry of Science and Higher Education of Poland under the international research fund 2018–2022 and the agreement number 3898/H2020/2018/2.

## APPENDIX. PASSING RATES IN DIFFERENT STUDY SIZES



**Fig. 7.** Passing rates for different fixed %ISRs as a function of the number of ISRs ( $n$ ) for study sizes  $N$  of 100 (a), 500 (b), and 5000 (c)

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

1. Shah VP, Midha KK, Dighe SV, McGilveray IJ, Skelly JP, Yacobi A, et al. Analytical methods validation: bioavailability,

- bioequivalence, and pharmacokinetic studies. *Pharm Res.* 1992;9(4):588–92. <https://doi.org/10.1023/A:1015829422034>.
2. Viswanathan CT, Bansal S, Booth B, DeStefano AJ, Rose MJ, Sailstad J, et al. Workshop/conference report - quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *AAPS J.* 2007;9(1):E30–42. <https://doi.org/10.1208/aapsj0901004>.
  3. Fast D, Kelley M, Viswanathan CT, O'Shaughnessy J, King S, Chaudhary A, et al. Workshop report and follow-up—AAPS workshop on current topics in GLP bioanalysis: assay reproducibility for incurred samples—implications of Crystal City recommendations. *AAPS J.* 2009;11(2):238–41. <https://doi.org/10.1208/s12248-009-9100-9>.
  4. Timmerman P, Luedtke S, van Amsterdam P, Brudny-Kloepfel M, Lausecker B. Incurred sample reproducibility: views and recommendations by the European Bioanalysis Forum. *Bioanalysis.* 2009;1(6):1049–56. <https://doi.org/10.4155/bio.09.108>.
  5. Guideline on bioanalytical method validation. Committee for Medicinal Products for Human Use. European Medicine Agency 2011. [https://www.ema.europa.eu/documents/scientific-guideline/guideline-bioanalytical-method-validation\\_en.pdf](https://www.ema.europa.eu/documents/scientific-guideline/guideline-bioanalytical-method-validation_en.pdf) Accessed 2018-11-21.
  6. Guidance Document. Conduct and analysis of comparative bioavailability studies health Canada. 2012. <https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/bioavailability-bioequivalence/conduct-analysis-comparative.pdf> Accessed 2018-11-21.
  7. Guidance for Industry Bioanalytical Method Validation. US Department of Health and Human Services, US Food and Drug Administration. 2018. <https://www.fda.gov/downloads/drugs/guidances/ucm070107.Pdf> Accessed 2018-11-21.
  8. Findlay JW, Kelley MM. ISR: background, evolution and implementation, with specific consideration for ligand-binding assays. *Bioanalysis.* 2014;6(3):393–402. <https://doi.org/10.4155/bio.13.339>.
  9. Fluhler E, Vazvaei F, Singhal P, Vinck P, Li W, Bhatt J, et al. Repeat analysis and incurred sample reanalysis: recommendation for best practices and harmonization from the Global Bioanalysis Consortium harmonization team. *AAPS J.* 2014;16(6):1167–74. <https://doi.org/10.1208/s12248-014-9644-1>.
  10. Subramaniam S, Patel D, Davit BM, Conner DP. Analysis of imprecision in incurred sample reanalysis for small molecules. *AAPS J.* 2015;17(1):206–15. <https://doi.org/10.1208/s12248-014-9689-1>.
  11. Rudzki PJ, Biecek P, Kaza M. Comprehensive graphical presentation of data from incurred sample reanalysis. *Bioanalysis.* 2017;9(12):947–56. <https://doi.org/10.4155/bio-2017-0038>.
  12. Rudzki PJ, Buś-Kwaśnik K, Kaza M. Incurred sample reanalysis (ISR): adjusted procedure for sample size calculation. *Bioanalysis.* 2017;9(21):1719–26. <https://doi.org/10.4155/bio-2017-0142>.
  13. Rudzki PJ, Kaza M, Biecek P. Extended 3D and 4D cumulative plots for evaluation of unmatched incurred sample reanalysis. *Bioanalysis.* 2018;10(3):153–62. <https://doi.org/10.4155/bio-2017-0210>.
  14. Kall MA, Michi M, van der Strate B, Freisleben A, Stoellner D, Timmerman P. Incurred sample reproducibility: 10 years of experiences: views and recommendations from the European Bioanalysis Forum. *Bioanalysis.* 2018;10(21):1723–32. <https://doi.org/10.4155/bio-2018-0194>.
  15. Arfvidsson C, Wilson A, Heijer M, Bailey C, Severin P, Milligan F, et al. Incurred sample reanalysis in AstraZeneca small molecule portfolio – what have we learned and where do we go next? *Bioanalysis.* 2018;10(21):1733–45. <https://doi.org/10.4155/bio-2018-0162>.
  16. Summerfield SG, Barfield M, White SA. Incurred sample reanalysis at GSK: what have we learned? *Bioanalysis.* 2018;10(21):1755–66. <https://doi.org/10.4155/bio-2018-0204>.
  17. Rocci ML Jr, Devanarayan V, Haughey D, Jardieu P. Confirmatory reanalysis of incurred bioanalytical samples. *AAPS J.* 2007;9(3):E336–43. <https://doi.org/10.1208/aapsj0903040>.
  18. Hoffman D. Statistical considerations for assessment of bioanalytical incurred sample reproducibility. *AAPS J.* 2009;11(3):570–80. <https://doi.org/10.1208/s12248-009-9134-z>.
  19. Thway TM, Eschenberg M, Calamba D, Macaraeg C, Ma M, DeSilva B. Assessment of incurred sample reanalysis for macromolecules to evaluate bioanalytical method robustness: effects from imprecision. *AAPS J.* 2011;13(2):291–8. <https://doi.org/10.1208/s12248-011-9271-z>.
  20. Voigtman E. Comparison of signal-to-noise ratios. *Anal Chem.* 1997;69(2):226–34. <https://doi.org/10.1021/ac960675d>.
  21. Sadygov RG, Yates JR. A hypergeometric probability model for protein identification and validation using tandem mass spectral data and protein sequence databases. *Anal Chem.* 2003;75(15):3792–8. <https://doi.org/10.1021/ac034157w>.
  22. Tabb DL, Fernando CG, Chambers MC. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res.* 2007;6(2):654–61. <https://doi.org/10.1021/pr0604054>.
  23. Kochen MA, Chambers MC, Holman JD, Nesvizhskii AI, Weintraub ST, Belisle JT, et al. Greazy: open-source software for automated phospholipid tandem mass spectrometry identification. *Anal Chem.* 2016;88(11):5733–41. <https://doi.org/10.1021/acs.analchem.6b00021>.
  24. Kaufmann A, Walker S, Mol G. Product ion isotopologue pattern: a tool to improve the reliability of elemental composition elucidations of unknown compounds in complex matrices. *Rapid Commun Mass Spectrom.* 2016;30(7):791–9. <https://doi.org/10.1002/rcm.7476>.
  25. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. <https://www.R-project.org> Accessed 2018-11-21.
  26. Color Brewer 2.0, color advice for cartography. <http://colorbrewer2.org/#type=diverging&scheme=RdYlBu&n=7> Accessed 2018-11-21.
  27. Draft Guidance for Industry Bioanalytical Method Validation. US Department of Health and Human Services, US Food and Drug Administration. 2013. <https://www.bioagilytix.com/wp-content/uploads/2016/02/FDA-Bioanalytical-Method-Validation-Draft-Guidance-2013.pdf> Accessed 2018-11-21.
  28. Diletti E, Hauschke D, Steinijans VW. Sample size determination for bioequivalence assessment by means of confidence intervals. *Int J Clin Pharmacol Ther Toxicol.* 1991;29(1):1–8.
  29. Challenge and Opportunity on the Critical Path to New Medicinal Products. US Food and Drug Administration. 2004. <http://wayback.archive-it.org/7993/20180125035500/https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/UCM113411.pdf>. Accessed 2018-11-21.
  30. Final endorsed Concept Paper M10: Bioanalytical Method Validation. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). 2016. [https://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Multidisciplinary/M10/ICH\\_M10\\_Concept\\_paper\\_final\\_7Oct2016.pdf](https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M10/ICH_M10_Concept_paper_final_7Oct2016.pdf). Accessed 2018-11-21