

Article

Artificial Intelligence Meets Bioequivalence: Using Generative Adversarial Networks for Smarter, Smaller Trials

Anastasios Nikolopoulos ¹  and Vangelis D. Karalis ^{1,2,*} 

¹ Department of Pharmacy, School of Health Sciences, National and Kapodistrian University of Athens, 15784 Athens, Greece; anastnik@pharm.uoa.gr

² Institute of Applied and Computational Mathematics, Foundation for Research and Technology Hellas (FORTH), 70013 Heraklion, Greece

* Correspondence: vkaralis@pharm.uoa.gr; Tel.: +30-210-727-4267

Abstract: This study introduces artificial intelligence as a powerful tool to transform bioequivalence (BE) trials. We apply advanced generative models, specifically Wasserstein Generative Adversarial Networks (WGANs), to create virtual subjects and reduce the need for real human participants in generic drug assessment. Although BE studies typically involve small sample sizes (usually 24 subjects), which may limit the use of AI-generated populations, our findings show that these models can successfully overcome this challenge. To show the utility of generative AI algorithms in BE testing, this study applied Monte Carlo simulations of 2×2 crossover BE trials, combined with WGANs. After training of the WGAN model, several scenarios were explored, including sample size, the proportion of subjects used for the synthesis of virtual subjects, and variabilities. The performance of the AI-synthesized populations was tested in two ways: (a) first, by assessing the similarity of the performance with the actual population, and (b) second, by evaluating the statistical power achieved, which aimed to be as high as that of the entire original population. The results demonstrated that WGANs could generate virtual populations with BE acceptance percentages and similarity levels that matched or exceeded those of the original population. This approach proved effective across various scenarios, enhancing BE study sample sizes, reducing costs, and accelerating trial durations. This study highlights the potential of WGANs to improve data augmentation and optimize subject recruitment in BE studies.



Academic Editor: Angelo Cangelosi

Received: 22 March 2025

Revised: 20 May 2025

Accepted: 21 May 2025

Published: 23 May 2025

Citation: Nikolopoulos, A.; Karalis, V.D. Artificial Intelligence Meets Bioequivalence: Using Generative Adversarial Networks for Smarter, Smaller Trials. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 47. <https://doi.org/10.3390/make7020047>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: artificial neural network; generative adversarial networks; generative AI algorithms; bioequivalence; sample size; clinical trials

1. Introduction

Sample size calculation is a crucial aspect of planning clinical trials, as it ensures that there are enough subjects to provide accurate and reliable assessments of drug products with a certain level of statistical assurance [1]. Appropriately sized samples are essential for making confident inferences that reflect the underlying population parameters. The required sample size to reject or accept a study hypothesis is determined by the statistical power of the utilized test. Studies with insufficient power often lead to unrealistic assumptions about treatment effectiveness, misjudgment of parameter variability, incorrect follow-up period estimates, the inability to predict compliance issues, high dropout rates, and failure to account for multiple endpoints. These errors result in wasted resources and potential harm to participants due to unwarranted expectations of therapeutic benefits. Adequate sample sizes and high-quality data collection contribute to more reliable, valid,

and generalizable findings, ultimately saving resources and ensuring ethical integrity in clinical research [2,3].

Similar issues arise in bioequivalence (BE) studies as in any clinical trial, particularly when comparing a generic pharmaceutical product (Test, T) with a reference product (R). These pharmaceutical products are considered bioequivalent if they contain the same active ingredient in the same molar dose, and this equivalence is confirmed through comparative pharmacokinetic studies, known as bioequivalence trials. When bioequivalence is demonstrated in these trials, the products can be regarded as therapeutically equivalent [4,5]. Determining an appropriate sample size for BE studies typically depends on five key study design parameters: the difference between the T and R formulation, the estimated residual variability, the desired statistical power (typically at least 80%), and the significance level (5%). In addition, when planning a BE study, the researcher should consider the possibility of dropouts and therefore plan to recruit an additional number of volunteers.

An in-depth analysis of BE studies underscores the pivotal role of statistics and computational methods. Pharmacokinetic modeling and simulation are indispensable in BE, often guiding significant regulatory guidelines [4,5]. The emergence of artificial intelligence (AI) has brought transformative capabilities to clinical research, using big data, machine learning algorithms, and neural networks to advance diagnosis, prognosis, monitoring, and treatment across diverse medical domains [6]. AI has notably revolutionized clinical trial automation, enhancing efficiency through the robust handling of extensive datasets and the generation of highly precise outcomes, and is increasingly integrated into healthcare, offering substantial benefits in patient monitoring, diagnosis, and treatment [7]. Modern AI techniques have significantly matured over the past five years, proving their potential to enhance clinical trial efficiency and patient care by optimizing early-stage trial processes and setting best practices [8,9]. AI has, also, the capability to identify patients who qualify for clinical trials, forecast which individuals are more likely to participate, and, at the same time, extract relevant features from electronic health records [10]. In medical imaging, AI technologies excel in analyzing complex data, such as distinguishing cell types in tumor samples with high accuracy, often surpassing traditional methods [11–13]. This capability extends to improving diagnostic precision and streamlining workflows through automated image analysis and predictive algorithms for medical emergencies [14]. However, integrating AI into healthcare presents challenges due to the industry's high regulation and risk aversion [8]. To address these, a stepwise approach is necessary, focusing on well-defined tasks where AI can offer clear improvements and thorough testing alongside existing technologies [15]. International guidelines like SPIRIT-AI and CONSORT-AI are crucial for ensuring a transparent and effective evaluation of AI systems in clinical trials, facilitating their safe and ethical use [16]. Despite the rapid development of AI, ongoing evaluation and adherence to best practices are essential to ensure its effective integration into clinical decision support [17–19]. While AI has the potential to revolutionize healthcare by enhancing trial efficiency, reducing costs, and improving patient care, achieving these benefits will require careful, incremental integration, rigorous evaluation, and robust regulatory frameworks [20–22]. Additionally, data augmentation techniques such as flipping, rotation, noise addition, shifting, cropping, PCA jittering, and Generative Adversarial Networks (GANs), including Wasserstein GANs (WGANs), play a crucial role. Notably, WGANs have demonstrated exceptional efficacy in this regard [20].

Our recent research explored the application of WGANs in clinical studies. This approach involved training a WGAN on a limited dataset, referred to as the “sample”, and then using it to generate virtual subjects that represent a larger population. Unlike traditional GANs, WGANs ensure stability and convergence through techniques such as weight clipping, which help the generated data to closely match real-world distributions.

In Monte Carlo simulations of clinical trials, we compared the performance of WGAN-generated virtual subjects with both the entire population (referred to as the “original” distribution) and a subset (the “sample”). The results demonstrated that WGANs can effectively create a virtual population from a small subset of data, producing outcomes comparable to those obtained from the entire population [21].

The aim of this study is to explore the applicability and performance of using WGANs as a tool to virtually increase the sample size in the assessment of generics, namely, the case of bioequivalence studies. In our previous study [21], we used very large original sample sizes (e.g., 10,000) to apply WGANs. However, in the current study, an intriguing challenge is that WGANs are applied to low sample sizes (e.g., 12, 24, 48, or 72), since bioequivalence studies are typically conducted with limited samples. Also, the conditions of most usually used 2×2 crossover trials are simulated in the current work and the statistical analysis follows the official requirements for BE trials, as imposed by the regulatory authorities (FDA, EMA). Additionally, several scenarios are explored to investigate as many aspects of BE as possible.

2. Materials and Methods

Evaluating the performance of WGANs in the context of BE studies requires replicating the conditions typical of such studies. The main components of the methodology included parameter selection, subject generation through Monte Carlo simulations, tuning WGAN hyperparameters, and sampling distributions. The samples were then used as input for WGANs to generate new distributions. All three distributions were evaluated for their BE acceptance according to regulatory standards. Each procedure was repeated hundreds of times to obtain robust results, varying the parameters each repetition (e.g., variability, T/R ratios, sample percentage, and original population size). The following sections provide a detailed explanation of each part.

2.1. Wasserstein Generative Adversarial Networks

GANs consist of two main parts: the generator and the discriminator. This architecture leads to a generative network through an adversarial process in which both neural networks are trained simultaneously. The generator creates new data instances that resemble the distribution of the original data, while the discriminator assesses the authenticity or quality of these generated instances, estimating the probability that a sample is either real or synthetic [22]. However, GANs are known for their high sensitivity to hyperparameters and can require a significant amount of time to converge. These training issues encountered with GANs led to the development of WGANs. WGANs use the Wasserstein loss as a meaningful distance measure in the loss function, effectively addressing the training failures associated with traditional GANs [22–24].

Figure 1 depicts the functioning of WGANs. Initially, the generator inputs random noise to produce an output, such as a value or an image. The critic then assesses both the real and generated outputs to differentiate between genuine and fake. The feedback from the critic is used to refine the generator, iteratively enhancing its capability to produce realistic outputs. The objective is for the generator to create outputs that are indistinguishable from real ones, so that the critic can no longer tell the difference between the two [25,26]. The inputs can be various objects, from images to tabular data [27].

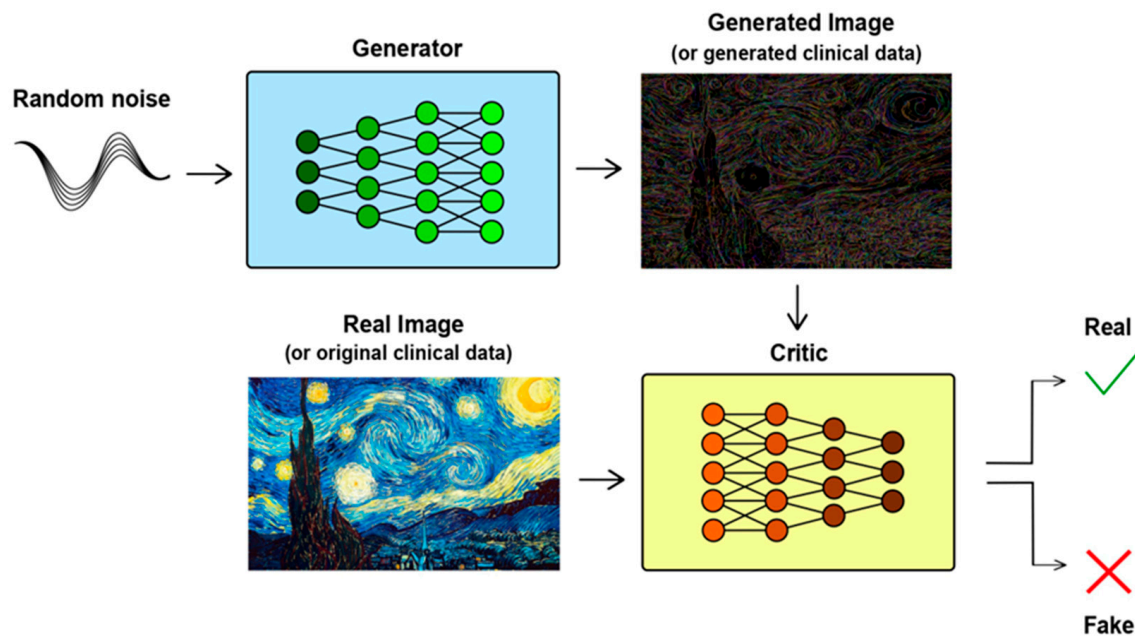


Figure 1. Visual representation of WGAN operation. This illustration simplifies the understanding of WGANs by using Vincent van Gogh’s famous painting ‘Starry Night’. The concepts depicted can be extended to clinical data scenarios and numerical data.

2.2. Hyperparameter Tuning

Hyperparameters play a crucial role in neural networks. Finding the optimal combination of epochs, latent dimensions, batch size, activation functions, number of hidden layers, number of neurons in hidden layers, and learning rates can be very challenging. However, this step is essential for the subsequent implementation of the algorithm. Table 1 lists the various hyperparameters evaluated during the algorithm fine-tuning process.

Table 1. Hyperparameters evaluated during the fine-tuning phase.

Epochs		250	500	750	1000
Latent dimensions		80	90	100	
Batch size		64	128		
Activation Function	Hidden Layers	tanh	linear		
	Output Layers	linear			
Number of Hidden Layers	Generator	1			
	Critic	1			
Number of Neurons in Hidden Layers	Generator	4	8	12	
	Critic	4	8	12	
Learning Rates		0.005–0.00005			

2.3. Monte Carlo Simulations

The simulations conducted in this study involved a 2×2 crossover design. In this design, subjects were randomly assigned to two groups: R and T, for both periods. These groups collectively constitute the population. For each chosen population size (N), half of the subjects were assigned to the R group and half to the T group. Subjects were generated with a mean endpoint (e.g., AUC) value of 100 units and a standard deviation determined by the mean and the between-subject coefficient of variability (CV). In the second period, the distributions were derived from the first period distributions, adjusted by a stochastic term to introduce within-subject variability (CVw). In this study, the mean CVw level was set to 10% and 25%, while CVw was allowed to vary between subjects.

After obtaining all four distributions (for the two groups (T, R) and the two periods of drug administration), random sampling was performed. The samples from these distributions were used as input for the WGANs' algorithm. The distributions generated by the WGANs constituted the generated population. All three distributions (population, sample, and generated population) underwent statistical analysis using the typical ANOVA model imposed by the regulatory authorities, namely, after $\ln$ -transformation of the values, etc. [4,5]. The residual error was then calculated to construct a 90% confidence interval for the mean difference between the two groups (T and R) in the $\ln$ -domain. BE acceptance was evaluated according to the regulatory framework for bioequivalence [4,5]. Each scenario was repeated 1000 times to calculate the BE acceptance percentage.

The flowchart depicted in Figure 2 outlines all the important steps of the research methodology.

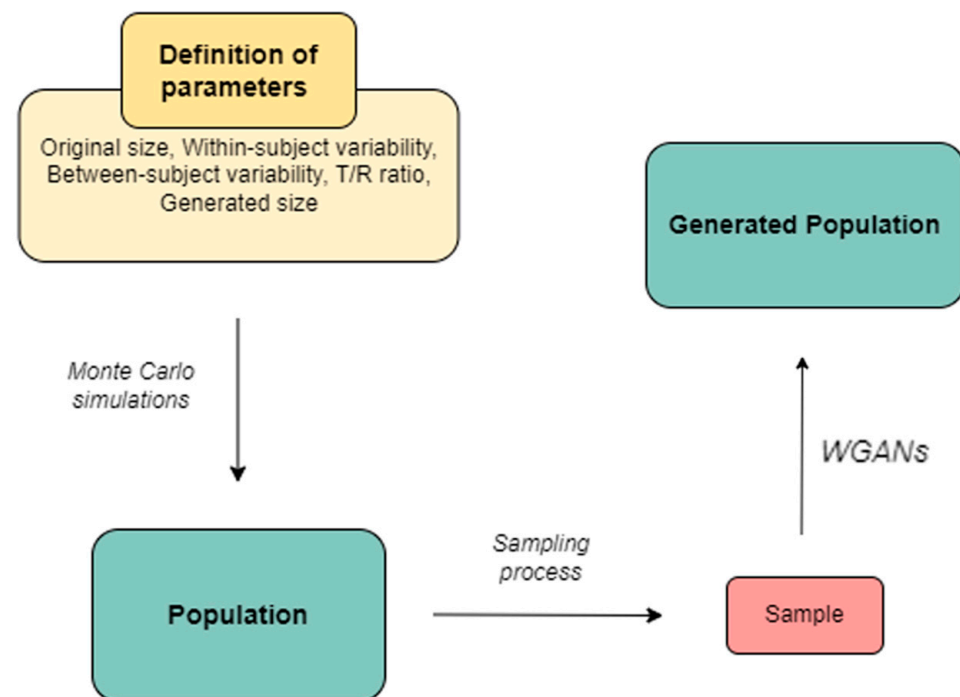


Figure 2. Flowchart depicting the methodology for implementing WGANs in bioequivalence studies. Initially, the entire population is created, followed by the sampling process, and ultimately applying WGANs to synthesize the generated population.

Our approach was based on the BE acceptance percentage of the three distributions and the similarity percentage. The similarity percentage was examined between (a) population vs. sample, and (b) population vs. generated subjects. This criterion refers to the simultaneous achievement of BE or not between the two compared distributions, for each different simulation. Providing another metric for the assessment of the WGANs was important, because without similarity, generated distributions can be biased, by either producing false negative results (masking true BE) or false positive results (increasing the BE acceptance without being true). Thus, by examining the similarity of the distributions, we can assess how much alike are the population and the sample (current approach), and the population and generated data (proposed approach).

Table 2 lists all the different parameter values for the scenarios that were used in the analysis. These scenarios represent all possible combinations of the values specified in the table.

Table 2. Scenarios for the Monte Carlo simulations of 2×2 crossover bioequivalence trials.

Test/Reference Ratios	Between-Subject Variability (CV)	Within-Subject Variability (CVw)	Size of the Population (N)	Sampling Percentage	Size of the Generated Distribution ($\times N$)
1.0	10%	10%	12, 24, 48, 72	25%	1 $\times$
1.05		25%		50%	2 $\times$
1.1				75%	

3. Results

Initially, we executed scenarios to evaluate BE acceptance and similarity percentages, maintaining an equal number of patients, N, in both the population and the generated population across various values of CVw, T/R ratios, and sampling percentages. The results were visualized through average histograms for each distribution: population, sample, and generated population. As illustrated in Figure 3, across three distinct scenarios, the generated distribution successfully captured values closely resembling those of the population. However, the relatively small N values pose a challenge in demonstrating the true resemblance of the distributions.

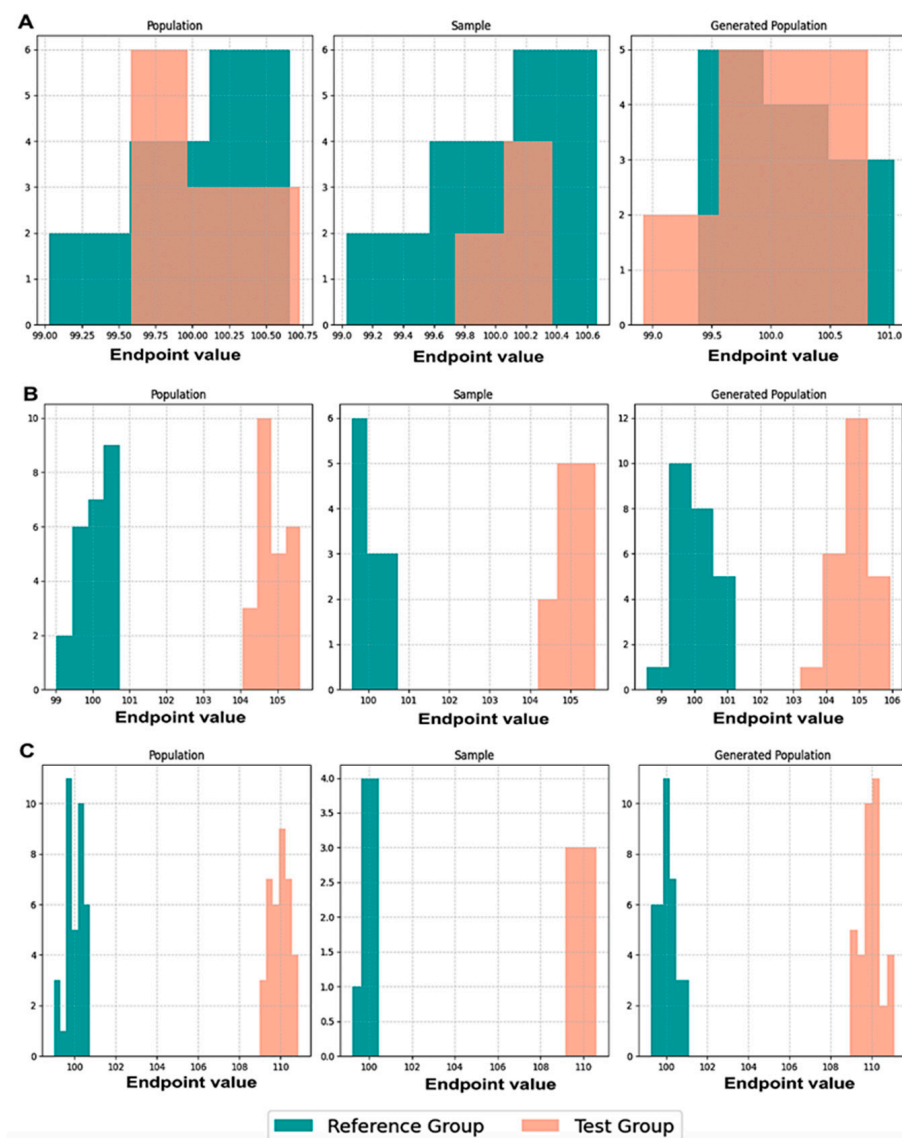


Figure 3. Average histograms for each distribution across three different scenarios: (A) $N = 24$, $T/R = 1.0$, 75% sampling, (B) $N = 48$, $T/R = 1.05$, 50% sampling, and (C) $N = 72$, $T/R = 1.1$, 25% sampling.

The true likeness of the population and generated distributions can further be seen in Figure A1, where the WGANs can resemble a population of 10,000 patients with just 50 of them as an input.

However, focusing on the scenarios where $N = 24$ for the results on the percentage of BE acceptance, it is evident that in all scenarios, across various values of CVw, T/R ratios, and sampling percentages, the generated population resembled the original population. In Figure 4, the WGANs were able to produce distributions that not only resembled the original population but also had identical BE acceptance percentages. To ensure accuracy, the values from the generated distributions were examined to avoid replication of the original distribution. This indicates that the WGANs created an entirely new distribution that mirrored the BE acceptance of the original, even when the input was as small as six patients.

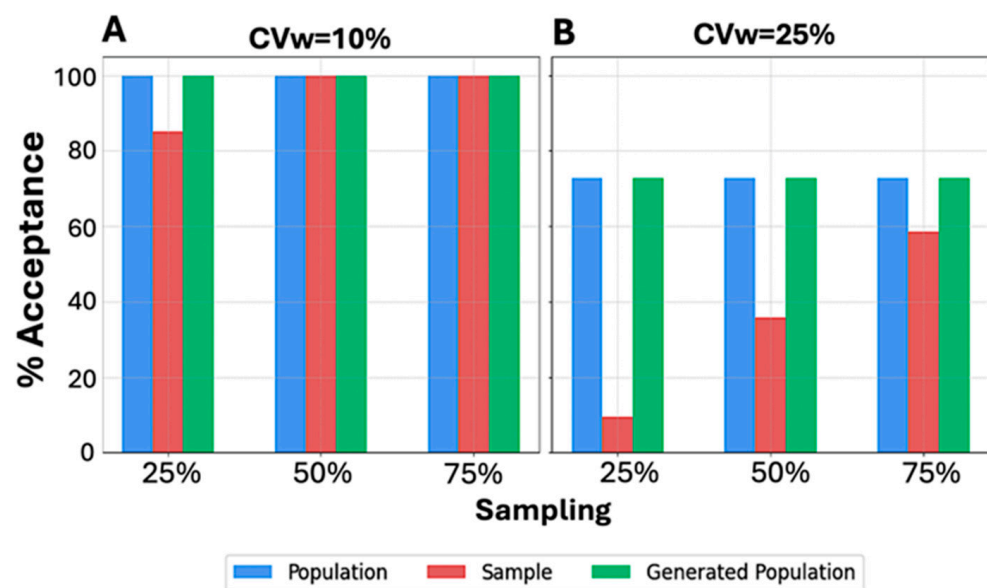


Figure 4. Percent bioequivalence acceptance for the population, the sample, and the generated subjects across various sampling percentages. In all cases, the population sample size was equal to 24, the T/R ratio equal to 1.05, and the residual variability (CVw) was set to 10% (A) and 25% (B).

A similar analysis was also performed (Figure 5) in the case of the other original sample size values, namely, 12, 48, and 72. A visual inspection of Figure 5A reveals that the BE acceptance percentages, specifically for CVw = 10%, exhibit the same pattern as in Figure 4. The BE acceptance percentages for the generated population consistently match those of the original population. This trend continues in all scenarios where CVw = 25%, as shown in Figure 5B. Despite the increased CVw, the WGANs consistently produced identical BE acceptance percentages to those of the original population, while the sample failed in most scenarios to resemble the original. Notably, the WGANs achieved statistical resemblance even in challenging scenarios, such as 25% sampling of $N = 12$, which corresponds to only three patients. Remarkably, only three patients were sufficient for the WGANs' algorithm to generate a distribution resembling the original, demonstrating the efficacy of this generative algorithm in very small sample sizes.

Focusing on scenarios where $N = 24$, Figure 6 illustrates the two types of similarity measured: (a) between the population and the sample (in red color), and (b) between the population and the generated subjects (in green color). It is worth mentioning that identical BE acceptance percentages resulted in 100% similarity across all scenarios, irrespective of the sampling percentage, CVw value, or T/R ratio (Figure 6).

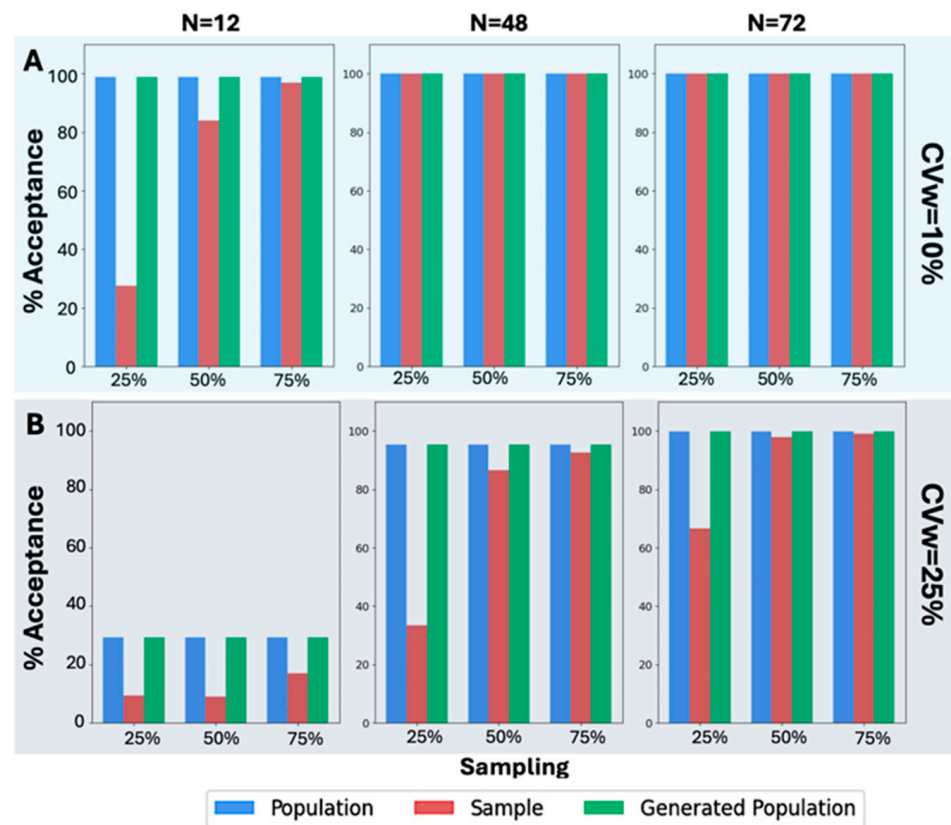


Figure 5. Percent bioequivalence acceptance for the population, the sample, and the generated subjects across various sampling percentages. Three population sample sizes were used: 12, 48, and 72, while the residual variability (CVw) was equal to 10% (A) and 25% (B).

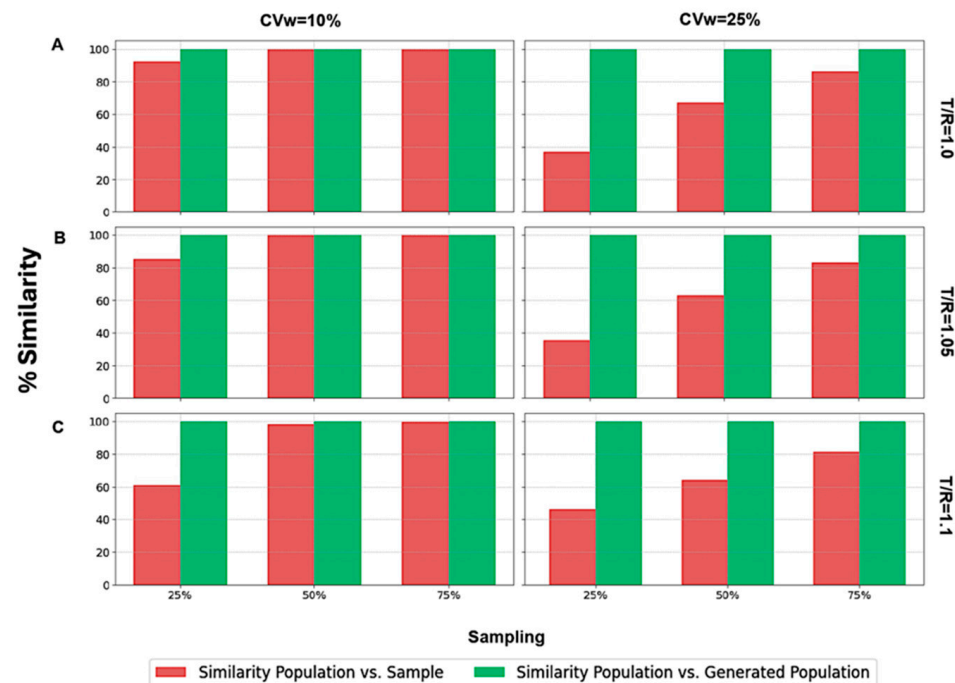


Figure 6. Similarity (%) between population sample (red) and between population-generated subjects (green) when N = 24 subjects is considered for the population. Two levels of residual variability (CVw) were explored: 10% and 25%. Panels (A), (B), and (C) correspond to T/R ratios of 1.0, 1.05, and 1.1, respectively.

On the contrary, only a few scenarios exhibited 100% similarity between the population and the sample. These instances occurred with a 75% sampling percentage and in two cases with a 50% sampling percentage (when $T/R = 1.0$ and $T/R = 1.05$). When the sampling percentage was as low as 25%, the sample consistently showed dissimilarity from the population, demonstrating that the sample did not resemble the population as closely as the generated population did. A similar analysis (Figure 7) was performed for the other three sample sizes (i.e., 12, 48, and 72).

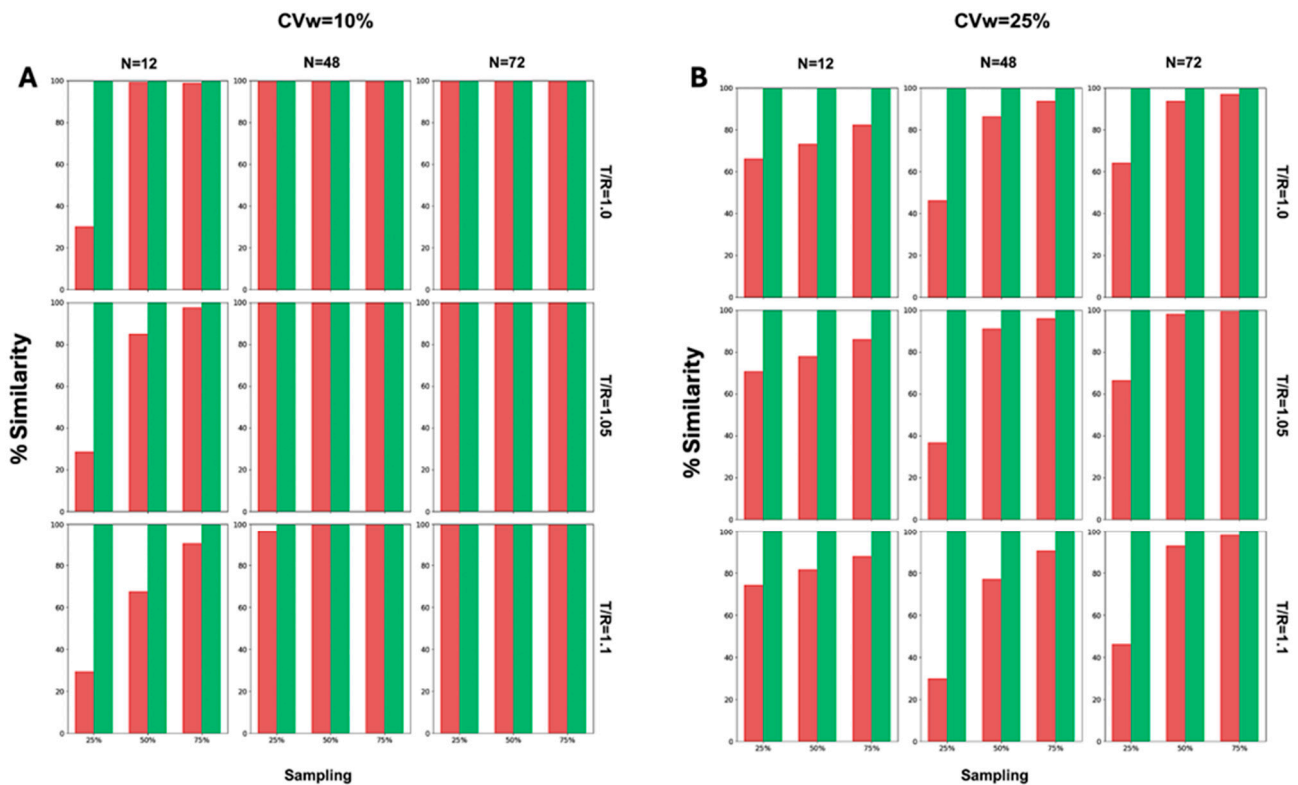


Figure 7. Similarity (%) between population sample (red) and between population-generated subjects (green) when the population consists of $N = 12, 48$, and 72 subjects. Two levels of residual variability (CVw) were explored: (A) $CVw = 10\%$ and (B) $CVw = 25\%$, while the horizontal panels refer to Test/Reference (T/R) ratios of $1.0, 1.05$, and 1.1 .

A visual inspection of Figure 7 shows that the same behavior was observed across all the scenarios examined. In particular, the similarity between the population and the generated subjects consistently was 100%, reflecting identical BE acceptance percentages. On the contrary, the similarity between the population and the sample was 100% only for large N values (namely, 48 and 72) and for sampling percentages of 50% and 75%. As the T/R ratio increased and the N value decreased, the sample became less similar to the population. In challenging scenarios, it is evident that only the generated population could accurately resemble the original population. The right part Figure 7 shows the similarity results for $CVw = 25\%$. In this case, no scenario achieved 100% similarity between the population and the sample, while WGANs consistently demonstrated 100% similarity with the population.

The final step of this analysis was to evaluate the performance of the WGANs in scenarios where the generated population was twice the size of the original population. Figure A2 illustrates the results for $N = 24$ across different T/R ratios, comparing BE acceptance percentages between the sample and the population, and between the generated population and the population. When the generated population was twice as large as the original, the WGANs exhibited a positive difference in BE acceptance, indicating that the

BE acceptance percentage of the generated population was equal to or greater than that of the original population. For CVw = 10% and T/R values of 1.0 and 1.05, there was no difference between the generated population and the original. However, for T/R = 1.1 and CVw = 25%, the difference was consistently positive. As T/R and CVw increased, the generated population showed improved BE acceptance percentages relative to the original population. Conversely, the difference between the sample and the population in terms of BE acceptance percentages was consistently negative, indicating that the sample had a lower acceptance percentage compared to the original population.

Continuing for the other N values, Figure A3 displays the same differences as Figure A2 but for the N values 12, 48, and 72, when CVw = 25%. Similar behavior is observed, with the difference between the BE acceptance of the generated population and the original population being consistently positive. As the number of patients decreases, the positive difference tends to increase, highlighting that WGANs perform better with smaller samples. Conversely, the sample consistently showed lower BE acceptance percentages compared to the population, suggesting that the sample may not adequately resemble the population. Both BE acceptance percentage and similarity are crucial for assessing the resemblance of distributions, for the scenarios where WGANs generated twice as many patients as the population. Figure A4 illustrates how the similarity percentage changes with increasing T/R ratios and CVw values for N = 12 and 75% sampling. For CVw = 10%, the similarity between the population and the generated population was consistently higher across all T/R ratios. However, for CVw = 25%, the similarity tended to decrease as the T/R ratio increased, remaining lower than the similarity between the population and the sample.

Finally, it is important to note again that the hyperparameters of the WGANs were kept constant across all scenarios to evaluate their performance. This consistency underscores the robustness of the WGAN architecture in handling diverse input data profiles, highlighting its adaptability and reliability across various clinical study settings.

4. Discussion

In our previous study, we introduced the idea of using WGANs in clinical studies to increase sample sizes by creating “virtual patients” [21]. WGANs, a variant of GANs, utilize the Wasserstein loss function to overcome the training challenges typically associated with GANs. Our study demonstrated that distributions generated by WGANs can effectively replicate the characteristics of the population, even with a sample size percentage as small as 0.5% of a population of 10,000. Moreover, it was found that traditional sampling methods could not accurately mirror the population, whereas the similarity between the generated population and the actual population was consistently greater than or equal to that of the sample to the population [21].

In our current study, a key challenge is the application of WGANs to significantly smaller sample sizes (e.g., 12, 24, 48, or 72), because BE studies are commonly conducted with limited participant numbers. To address this, we simulate the conditions typically seen in 2×2 crossover trials, the most frequently used design for BE studies. Furthermore, our statistical analyses are rigorously aligned with the official requirements set forth by regulatory authorities, such as the FDA and EMA, ensuring compliance with the established standards for BE trials. This approach allows us to evaluate the robustness of WGANs under realistic BE study constraints, potentially advancing their application in regulatory science. In this work a comparison is made among the population, sample, and generated subjects in two ways: (a) BE acceptance (statistical power) and (b) similarity with the performance of the actual population. The ultimate purpose of this study is to show how the incorporation of data augmentation techniques, such as generative AI algorithms, in

clinical research can minimize costs and study durations, as well as reduce human exposure by substituting part of the actual volunteers with “virtual subjects”.

In bioequivalence studies, regulatory agencies typically require at least 80% statistical power to detect true differences between formulations. Meeting this standard often necessitates large sample sizes, driving up costs, extending timelines, and raising ethical concerns due to increased drug exposure. A promising alternative, supported by advances in generative AI, is the use of AI-synthesized virtual subjects that replicate human pharmacokinetic behavior. The aim is not to exceed the 80% power threshold but to maintain it while reducing the need for real participants. By integrating virtual subjects into study designs, researchers can minimize human exposure, lower costs, and accelerate timelines—without compromising statistical or regulatory standards.

It is important to clarify that the purpose of incorporating AI-synthesized virtual subjects is to partially substitute, not entirely replace, the need for human participants in bioequivalence or clinical studies. Our proposed approach involves enrolling a limited number of human subjects and using their pharmacokinetic data to train a generative adversarial network, specifically a Wasserstein GAN, to simulate virtual counterparts. This method is designed for application in clinical and BE trials, where regulatory power requirements must be met efficiently. It should be emphasized that this approach is not intended for use during the earlier stages of drug development, such as formulation design or lot selection. In those contexts, other modeling techniques are more appropriate [28]. For instance, in generic drug development, the In Vitro–In Vivo Simulation (IVIVS) approach has been proposed as a semiphysiological tool [29]. IVIVS can guide R&D efforts by linking dissolution properties to expected in vivo performance, enabling predictions of bioequivalence outcomes based solely on in vitro dissolution data. Nevertheless, formulation optimization conducted with a limited number of subjects during the early stages of product development may generate sufficient PK data to support the subsequent application of a WGAN model, which, once implemented, could facilitate the progression of the development program.

Methodologically, Monte Carlo simulations were applied to generate subjects categorized into Test or Reference groups. Random sampling was conducted at 25%, 50%, and 75% levels, with the sampled subjects serving as input for the WGANs. To ensure robust results, this procedure was repeated 1000 times for each scenario, with varying parameters such as within-subject variability (i.e., the CV_w), the ratio of average endpoint values between Test and Reference groups (i.e., the T/R ratio), sampling percentages, and the size of the generated distributions (either 1× or 2× the original size).

In our study, the WGAN loss function and the similarity criterion served distinct but complementary roles. The WGAN loss function was used during the training phase to ensure stable and effective learning of the generative model. By minimizing the Wasserstein distance between real and generated data distributions, it provided smoother gradients and mitigated issues such as mode collapse, which is especially critical in BE trials that typically involve small sample sizes. This loss function guided the generator in producing virtual subjects whose statistical properties closely mirrored those of actual participants. In contrast, the similarity criterion was applied after training as an evaluation measure to determine how well the synthetic populations replicated the functional outcomes of real BE trials. Specifically, it involved comparing the results of Monte Carlo–simulated 2 × 2 crossover BE studies using real and generated subjects, focusing on the simultaneous BE acceptance of PK parameters like AUC and C_{max}. Rather than merely assessing distributional similarity, this criterion evaluated whether the AI-generated populations could achieve comparable regulatory decisions, thereby validating their utility in BE assessments.

The exhaustive approach followed in this work aimed to replicate real-life conditions while evaluating the WGANs' performance in challenging scenarios. A T/R ratio of 1.0 indicates identical performance between the groups, resulting in a relatively high BE acceptance percentage. Conversely, a larger T/R ratio, such as 1.1, indicates a 10% average difference between the groups, leading to reduced BE acceptance percentages. It was selected CVw values of 10% (small variability) and 25% (moderate variability) to test the algorithm under different conditions. Our results showed that WGANs performed well even in scenarios with relatively high T/R ratios (i.e., $T/R = 1.1$) or increased CVw (25%), demonstrating their capability to generate distributions like the original under challenging conditions. However, further studies on drugs with high variability are needed to evaluate WGANs' performance in extreme scenarios. Lastly, extensive hyperparameter tuning was conducted to optimize the WGANs' performance.

Across all scenarios where the size of the generated subjects matched the size of the original population, the generated distributions consistently exhibited identical BE acceptance percentages to those of the original distributions. This consistency is shown in Figures 4 and 5, which show that even in challenging scenarios, the generated subjects closely resembled the original ones. It is also worth mentioning that this performance was consistent in all cases regardless of the original population size (i.e., the N).

Moreover, the abovementioned attribute is translated into a 100% similarity between the original population and the generated subjects. Figures 6 and 7 further illustrate this point. The similarity percentages reveal that the samples could not match the original population as effectively as the generated subjects, emphasizing the potential risks of using small samples in BE studies. Notably, WGANs were able to generate distributions resembling the original with just a 25% sampling percentage. This underscores the capability of WGANs to work well with small sample sizes, capture the underlying statistical characteristics, and generate distributions that closely mirror the originals.

Regarding scenarios where the generated population was twice the size of the original, the WGAN-generated distributions achieved either a positive or zero difference in BE acceptance percentage compared to the original distribution (Figures A2 and A3). This indicates that the generated population consistently had equal or higher BE acceptance percentages compared to the original population. Conversely, the sample consistently showed a negative or zero difference in BE acceptance percentages compared to the original distribution. WGANs demonstrated superior performance in scenarios with smaller N values, regardless of the T/R ratio, CVw, or sampling percentages, by achieving a higher percentage of BE acceptance. When comparing similarities (Figure A4), the similarity between the population and the generated population tended to decrease as the T/R ratio and CVw increased to 25%. When CVw was 10%, the similarity between the population and the generated population was consistently higher in all scenarios compared to the similarity between the population and the sample. Conversely, when CVw was 25%, the similarity between the population and the generated population decreased as the size of the population, N, decreased.

In our WGAN-based framework for generating virtual patients in BE studies, the primary input to the model is a sampled set of PK data derived from simulated 2×2 crossover BE studies. Our simulations focused on a single PK parameter (i.e., AUC, Cmax) under the common assumption of a log-normal distribution. To mimic the conditions of actual BE trials, inter-subject and within-subject variability were incorporated into the simulations. The WGAN outputs a synthetic population of virtual patients whose PK profiles preserve the key statistical characteristics of the original sampled population. The model learns distributional patterns in the training data, including variability and correlations, and generates new, realistic data points suitable for further BE simulations.

Key features learned by the WGAN include the shape, scale, and variability structure of the PK parameter within the simulated BE design. As a result, the generated virtual patients exhibit pharmacokinetic behavior that is statistically consistent with that of real subjects under similar trial conditions, supporting meaningful BE analysis based on synthetic data. One of the advantages of using WGANs in our approach is that there are no restrictions on the type of input data of the model, because the output will reflect the input distribution. These data can be pharmacokinetic or a variety of different endpoints, thus enhancing the applicability of our approach in different types of data.

Since in this study we proposed a new idea in the field, it is important to avoid any misunderstanding. The proposed WGAN method is designed to be applied directly to a limited, underpowered sample of actual human subjects from a specific clinical/BE study. The key idea is to use the available data from this small sample to train the WGAN, which then synthesizes virtual subjects that reflect the observed variability and structure of the real trial participants. We deliberately avoid using external sources of population PK parameters, as doing so would be both impractical and potentially introduce significant bias. This is because the WGAN is intended for use in the context of BE studies, which involve a head-to-head comparison of two specific pharmaceutical products (typically a reference and a test formulation). While some PK data might exist for the reference product, such information is typically unavailable for the test product, especially when it is a new or generic formulation. Thus, drawing from external data sources or the literature (including Bayesian methods) would not accurately represent the unique characteristics of the test product under investigation. Importantly, the objective of this approach is not to reconstruct missing data for a known drug but rather to model the comparative performance of two formulations within the specific context of the BE trial. Relying on external data would undermine this aim and compromise the integrity of the virtual population.

It is important to address some limitations of this research. The execution time for each scenario required a significant amount of time and substantial computational resources. While more advanced systems could provide faster training and quicker results, the computational resources available to us were sufficient to successfully execute the scenarios in our research. In bioequivalence studies, while WGANs can improve the quality of data generation and support more robust statistical analyses, careful training of the model is essential to avoid biases that may be amplified during data generation [30]. This is particularly critical when assessing drug side effects. Additionally, the importance of generating “virtual patients” from a limited pool of real patient data should not be overlooked.

In addition, to further advance the application of AI-generated subjects in BE studies, it is essential to expand the current framework beyond low-to-moderate variability cases (e.g., 10% and 25%) and explore scenarios involving highly variable drugs. These drugs, often characterized by an intra-subject variability exceeding 30%, present significant challenges in BE assessment and require more sophisticated modeling of distributional extremes, including outliers and skewed data. Additionally, future research should examine cases where the GMR exceeds the conventional BE acceptance limits (e.g., $GMR > 1.25$) to explicitly evaluate the model behavior under non-bioequivalent conditions and assess type II error rates. This will help determine whether the virtual population approach maintains appropriate sensitivity and specificity in detecting true differences between formulations.

It is important to note that while our approach increases statistical power by augmenting datasets with “virtual patients”, the evidence for demonstrating BE must continue to originate from real human subjects. The integrity of BE assessments depends on data derived from actual clinical trial participants, and this remains a non-negotiable standard for regulatory acceptance. In this context, it is also essential that the enrolled human

subjects constitute a representative sample, in accordance with general statistical sampling principles, to ensure the external validity and applicability of the study findings. Our method is not intended to replace human-based evidence, but rather to enhance the efficiency, robustness, and ethical conduct of BE studies. By incorporating AI-generated virtual subjects, researchers can reduce the number of real participants required, lower study costs, minimize unnecessary human exposure, and shorten timelines—while still meeting statistical and regulatory standards.

Given that generated data must undergo extensive validation, regulatory acceptance remains limited, and the applicability of such algorithms must be carefully examined in real-world scenarios. This study is the first to introduce the use of WGANs in BE research, offering a foundational proof-of-concept for leveraging generative AI to create virtual patients within the context of generic drug evaluation. While our approach generates virtual subjects based on the distributional characteristics of the underlying population, we acknowledge that real-world BE studies often present complexities such as extreme values, outliers, and deviations from lognormality. This issue can be addressed in several ways. For example, in the study of Burger et al., it was proposed a robust Bayesian method using skew-t distributions to assess the average BE in crossover studies [31]. The proposed approach effectively handled outliers and skewed data to improve inference accuracy and statistical power. Thus, a possible solution could be the incorporation of similar Bayesian methods in the WGAN model. A similar approach has recently been proposed in the form of BayesT, which is a non-informative Bayesian method designed to assess BE as an alternative to the traditional Two One-Sided t-Tests. BayesT offers greater robustness in datasets with deviations from lognormality or extreme values, where it showed more consistent and potentially less conservative results. Additionally, its ability to be pre-specified for regulatory use provides a structured and reliable alternative for evaluating BE in challenging scenarios [32].

Another way is to rely on the WGAN itself, because it is feasible to train a WGAN model on data that deviate from normality or lognormality. WGANs are particularly well suited for modeling complex, non-Gaussian distributions due to their use of the Wasserstein distance, which provides more stable training dynamics and meaningful loss metrics than traditional GANs [33]. However, outliers can negatively impact training by destabilizing the critic, distorting gradient updates, and biasing the generator toward anomalous regions of the data space. To address these effects, preprocessing steps such as normalization, robust scaling, or outlier-handling techniques (e.g., clipping or statistical detection) can be used. Additionally, using the gradient penalty variant (WGAN-GP) instead of weight clipping has been shown to improve convergence and stability, especially in datasets with heavy-tailed distributions or significant outlier presence [34–38]. These measures help ensure the model remains robust and capable of generating high-quality synthetic data that accurately reflect the underlying distribution. Nevertheless, it should be mentioned that this study represents an initial step in a broader and rapidly evolving field. Substantial work remains, including ongoing and future efforts to validate these approaches—particularly in addressing non-normal distributions and the impact of outliers. Progress in this area will also require close collaboration with regulatory bodies to define clear, acceptable pathways for integrating AI-generated data into regulatory decision-making. In parallel, related challenges such as high variability are being investigated by our research team [39].

By generating synthetic data based on existing datasets, AI-driven augmentation methods expand sample sizes, improve model performance, and help mitigate biases in the data. Among these, the implementation of diffusion models has recently emerged. For example, these models have recently been shown to enhance both performance and fairness in tabular data tasks by generating synthetic samples that better capture the variability of

underrepresented groups [40]. Similarly, in natural language processing, AI-based augmentation techniques including paraphrasing and back-translation have been used to increase data diversity and reduce model bias, particularly in low-resource scenarios [41]. Finally, Variational Autoencoders have been thoroughly examined in different scenarios in the field of clinical and BE studies, highlighting their advantages in recent research [42]. These developments illustrate the broader applicability of AI in augmenting datasets to improve robustness and generalizability [43]. In the context of BE and pharmacokinetic studies, such approaches offer an opportunity to simulate larger and more diverse virtual populations, which can enhance the statistical power of studies without additional human recruitment.

The aim of this study was to introduce the concept of using generative AI algorithms in the assessment of generics, where limited numbers of healthy volunteers are utilized. This is a major challenge in clinical research, as we aim to minimize actual human exposure in clinical trials while still achieving the desired statistical power and ensuring that the study performance (with AI-synthesized volunteers) matches that of a study with a large sample size. Our proposed approach does not intend to suggest that current regulatory standards call for higher BE power, but rather to highlight that increasing power beyond the standard threshold can offer practical benefits, such as reducing sample sizes, improving the robustness of the study, and addressing challenging cases (e.g., extreme scenarios of highly variable drugs). Our research aims to enhance study efficiency while remaining fully compliant with existing regulatory standards. The ultimate purpose is thereby reducing the need for human exposure in BE trials. Generative AI algorithms, such as WGANs, offer significant advantages in generating distributions that closely resemble the original population. These algorithms achieve better performance with small sample sizes, whereas traditional samples often fail to capture the population characteristics.

It should be emphasized that this study is the first to introduce the use of WGANs in BE assessment. This study represents only a novel and early-stage concept in applying artificial intelligence to regulatory science. As such, we anticipate that initial regulatory feedback will focus on understanding the foundational principles of this approach, particularly the processes involved in generating and validating virtual patient data. Regulatory agencies will require time to fully evaluate the methodology and its implications and will likely seek a clearly defined, step-by-step framework detailing how generative models are trained, validated, and applied in BE contexts. A key area of regulatory interest will be the reproducibility of results, namely, whether AI-generated virtual populations can consistently produce trial outcomes that align with those derived from real human subjects. Moreover, regulators will expect robust evidence demonstrating that synthetic data can preserve BE acceptance decisions across a range of study scenarios, including those with real-world complexities such as outliers, non-normal distributions, and high inter-subject variability. For this new approach to gain regulatory acceptance, it must be accompanied by transparent methodology, rigorous statistical validation, and a clearly articulated roadmap for integration into current regulatory pathways. Our future work will focus on collaborating with regulatory authorities to co-develop such frameworks and address core concerns related to scientific rigor, ethical standards, and practical implementation.

5. Conclusions

In a previous study, we introduced the use of artificial neural networks, particularly WGANs, in clinical research as a tool to reduce the number of required human subjects. In the present study, we extend this concept to bioequivalence trials, specifically focusing on the assessment of generic drugs. A key challenge in this context lies in the fact that bioequivalence studies are typically conducted with limited sample sizes (usually around 24 subjects), which may restrict the applicability of AI-synthesized (“virtual”) subjects. To

evaluate the performance of the proposed approach, we conducted Monte Carlo simulations of 2×2 crossover bioequivalence studies, incorporating WGANs. This allowed us to examine a wide range of scenarios (e.g., several levels of variability, sample sizes, and sampling percentages) that are commonly encountered in bioequivalence assessments. It should be underlined that the AI-generated populations consistently demonstrated bioequivalence acceptance rates comparable to, or even exceeding, those of the original populations, while traditional sampling methods failed to replicate the population characteristics. These findings suggest that AI-driven generative algorithms can deliver similar or superior outcomes with fewer real data points. Our research supports the integration of AI-based methods in clinical trials. Nevertheless, for their successful adoption, regulatory agencies will need to establish clear guidelines and criteria to govern the use of AI-generated virtual subjects, ensuring consistency and mitigating potential risks.

Author Contributions: Conceptualization, V.D.K.; methodology, V.D.K.; software, A.N.; validation, A.N.; formal analysis, A.N.; investigation, A.N.; resources, V.D.K.; data curation, A.N.; writing—original draft preparation, A.N.; writing—review and editing, V.D.K.; visualization, A.N. and V.D.K.; supervision, V.D.K.; project administration, V.D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Raw data were generated through simulations. Derived data supporting the findings of this study are available from the corresponding author [V.D.K.] on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

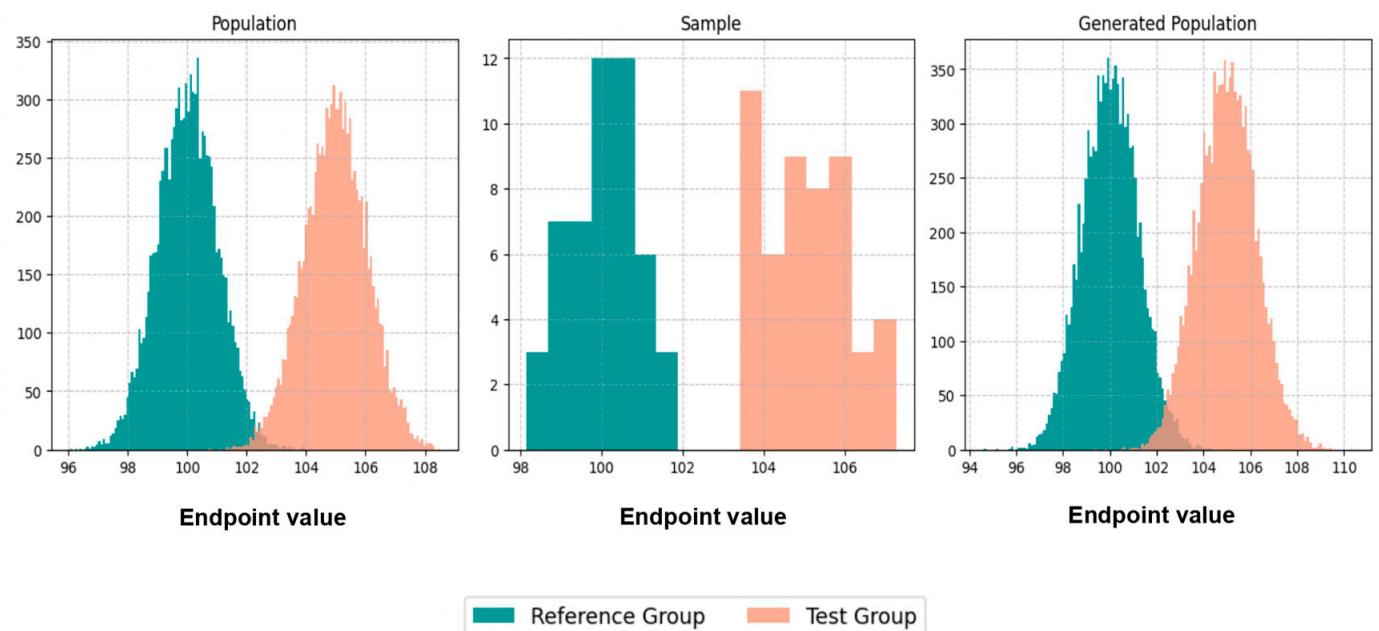


Figure A1. Average histograms of each distribution for an original population sample of $N = 10,000$ subjects, Test/Reference = 1.05, and between-subject variability of 10%.

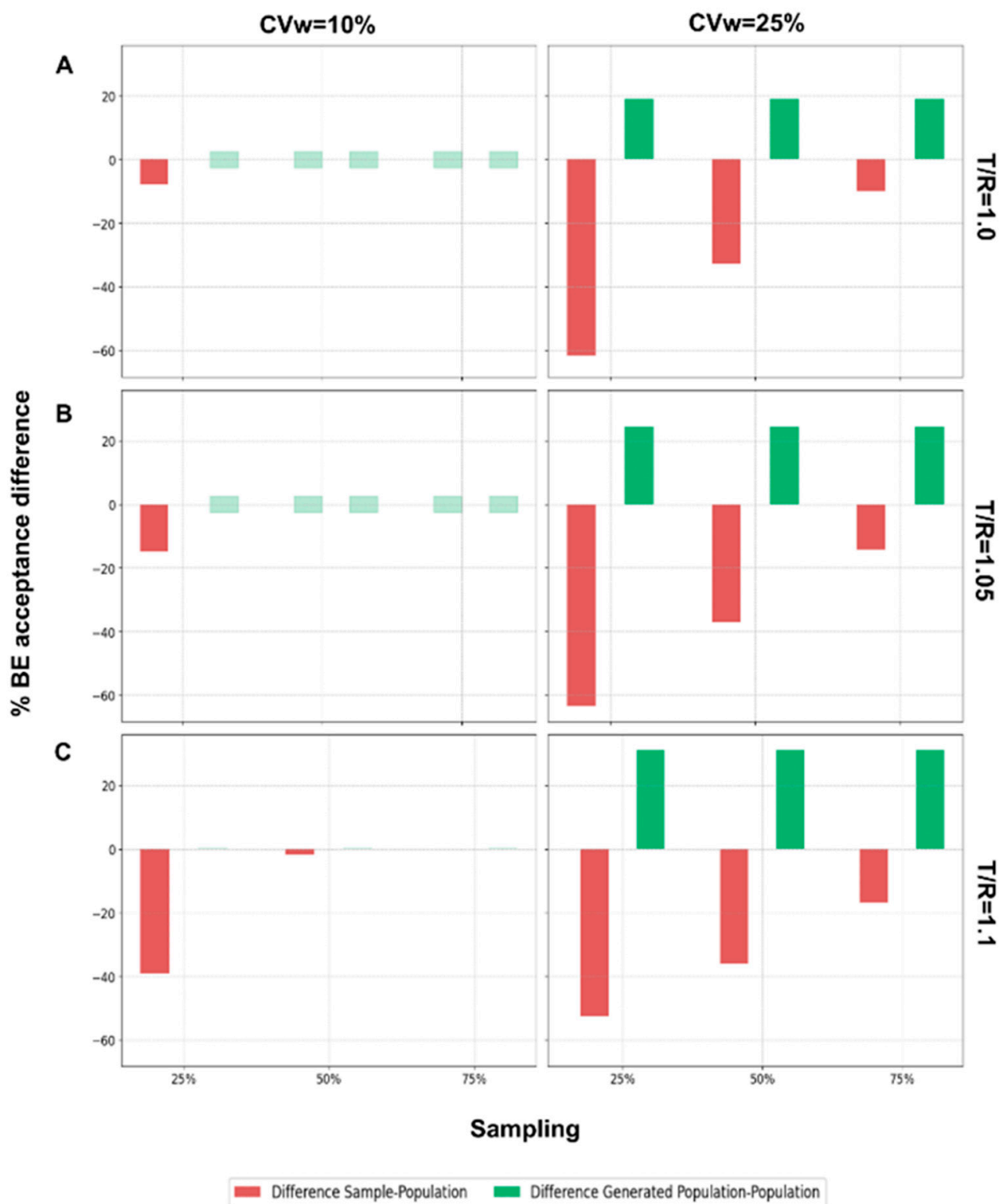


Figure A2. The bars represent the positive or negative difference in bioequivalence acceptance percentages for $N = 24$, across various sampling percentages. Panels (A), (B), and (C) correspond to Test/Reference (T/R) ratios of 1.0, 1.05, and 1.1, respectively. The light green rectangles (in panels (A,B)) indicate zero difference.

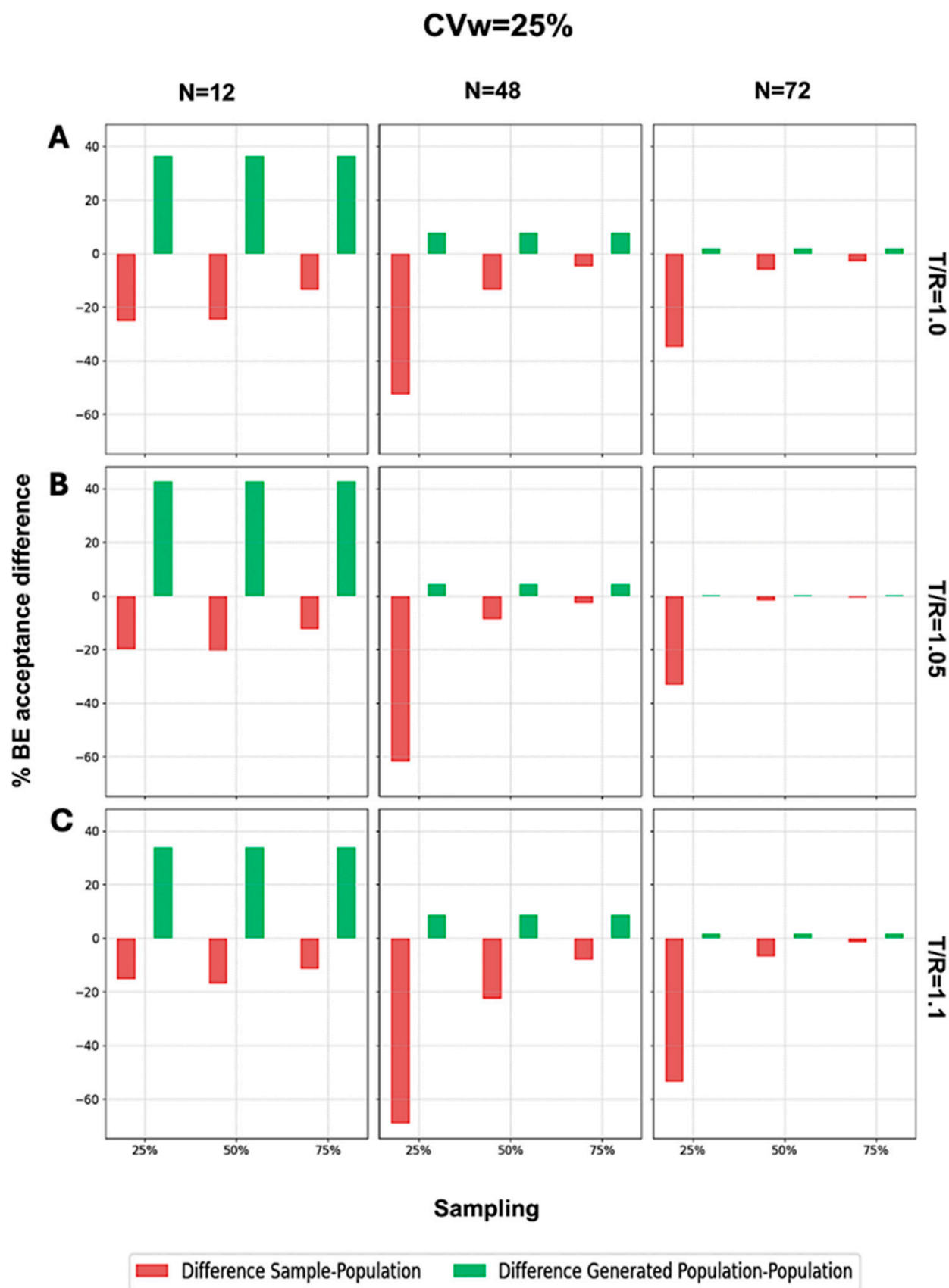


Figure A3. Positive (green bars) or negative (red bars) difference in bioequivalence acceptance percentages for $N = 12, 48, 72$, and residual variability (CVw) equal to 25% across various sampling percentages. Panels (A), (B), and (C) correspond to Test/Reference (T/R) ratios of 1.0, 1.05, and 1.1, respectively.

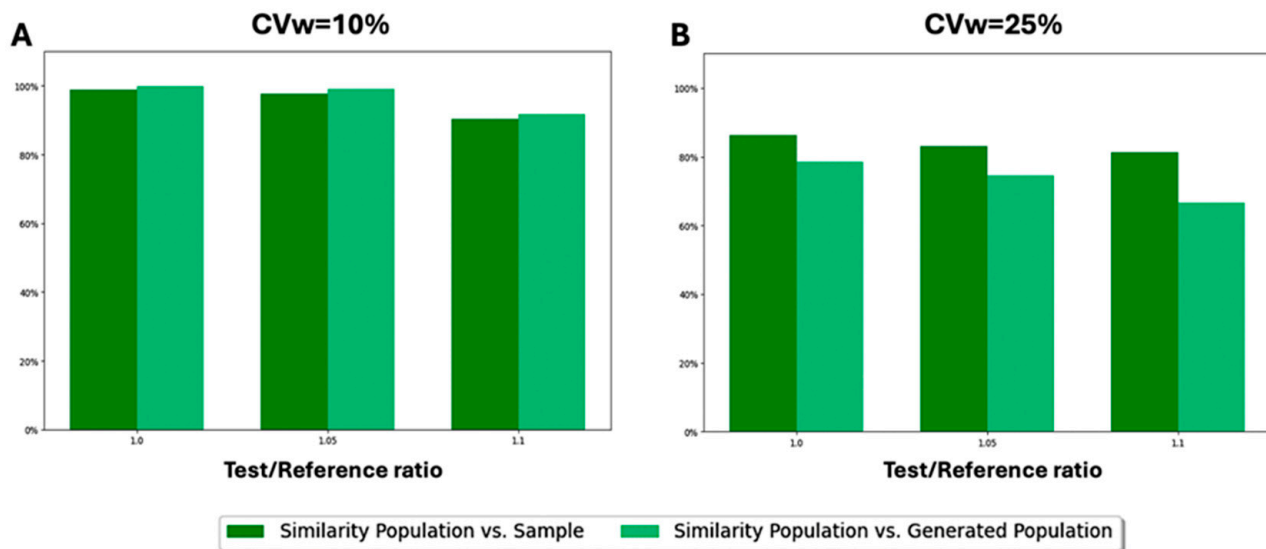


Figure A4. The bars represent the decrease in similarity percentages for N = 12 and 75% sampling, across different Test/Reference ratios and residual variability (CVw) values. Two types of bars are shown depending on the similarity between the following: (A) population vs. sample (intense green color), and (B) population vs. generated subjects (light green color).

References

1. Chow, S.C.; Shao, J.; Wang, H.; Lokhnygina, Y. *Sample Size Calculations in Clinical Research*, 3rd ed.; Taylor & Francis: Boca Raton, FL, USA, 2017. [CrossRef]
2. Suresh, K.; Chandrashekar, S. Sample Size Estimation and Power Analysis for Clinical Research Studies. *J. Hum. Reprod. Sci.* **2012**, *5*, 7–13. [CrossRef]
3. Eng, J. Sample size estimation: How many individuals should be studied? *Radiology* **2003**, *227*, 309–313. [CrossRef]
4. European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP). Guideline on the Investigation of Bioequivalence. CPMP/EWP/QWP/1401/98 Rev. 1/Corr 2010. Available online: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf (accessed on 15 January 2025).
5. U.S. Food and Drug Administration (FDA). Guidance for Industry: Bioavailability and Bioequivalence Studies Submitted in NDAs or INDs—General Considerations. Draft Guidance, December 2013. Available online: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioavailability-and-bioequivalence-studies-submitted-ndas-or-ind-general-considerations> (accessed on 15 January 2025).
6. Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Future Hosp. J.* **2019**, *6*, 94–98. [CrossRef] [PubMed]
7. Busnatu, Ș.; Niculescu, A.G.; Bolocan, A.; Petrescu, G.E.D.; Păduraru, D.N.; Năstasă, I.; Lupușoru, M.; Geantă, M.; Andronic, O.; Grumezescu, A.M. Clinical applications of artificial intelligence—An updated overview. *J. Clin. Med.* **2022**, *11*, 2265. [CrossRef] [PubMed]
8. Van Hartskamp, M.; Consoli, S.; Verhaegh, W.; Petkovic, M.; Van De Stolpe, A. Artificial Intelligence in Clinical Healthcare Applications: Viewpoint. *Interact. J. Med. Res.* **2019**, *8*, e12100. [CrossRef] [PubMed]
9. Rajpurkar, P.; Chen, E.; Banerjee, O.; Topol, E.J. AI in Health and Medicine. *Nat. Med.* **2022**, *28*, 31–38. [CrossRef]
10. Liu, R.; Rizzo, S.; Whipple, S.; Pal, N.; Pineda, A.L.; Lu, M.; Arnieri, B.; Lu, Y.; Capra, W.; Copping, R.; et al. Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature* **2021**, *592*, 629–633. [CrossRef] [PubMed]
11. Haleem, A.; Javaid, M.; Khan, I.H. Current status and applications of artificial intelligence (AI) in the medical field: An overview. *Curr. Med. Res. Pract.* **2019**, *9*, 231–237. [CrossRef]
12. Motamed, S.; Rogalla, P.; Khalvati, F. Data Augmentation Using Generative Adversarial Networks (GANs) for GAN-Based Detection of Pneumonia and COVID-19 in Chest X-ray Images. *Inform. Med. Unlocked* **2021**, *27*, 100779. [CrossRef]
13. Castiglioni, I.; Rundo, L.; Codari, M.; Di Leo, G.; Salvatore, C.; Interlenghi, M.; Gallivanone, F.; Cozzi, A.; D’Amico, N.C.; Sardanelli, F. AI applications to medical images: From machine learning to deep learning. *Phys. Med.* **2021**, *83*, 9–24. [CrossRef]
14. Harrer, S.; Shah, P.; Antony, B.; Hu, J. Artificial intelligence for clinical trial design. *Trends Pharmacol. Sci.* **2019**, *40*, 577–591. [CrossRef]

15. Ibrahim, H.; Liu, X.; Rivera, S.C.; Moher, D.; Chan, A.W.; Sydes, M.R.; Calvert, M.J.; Denniston, A.K. Reporting guidelines for clinical trials of artificial intelligence interventions: The SPIRIT-AI and CONSORT-AI guidelines. *Trials* **2021**, *22*, 11. [\[CrossRef\]](#)
16. Magrabi, F.; Ammenwerth, E.; McNair, J.B.; De Keizer, N.F.; Hyppönen, H.; Nykänen, P.; Rigby, M.; Scott, P.J.; Vehko, T.; Wong, Z.S.; et al. Artificial intelligence in clinical decision support: Challenges for evaluating AI and practical implications. *Yearb. Med. Inform.* **2019**, *28*, 128–134. [\[CrossRef\]](#)
17. Shaheen, M.Y. Applications of Artificial Intelligence (AI) in Healthcare: A Review. *Sci. Prepr.* **2021**. [\[CrossRef\]](#)
18. Komorowski, M.; Celi, L.A.; Badawi, O.; Gordon, A.C.; Faisal, A.A. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat. Med.* **2018**, *24*, 1716–1720. [\[CrossRef\]](#)
19. Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V.I. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 310. [\[CrossRef\]](#)
20. Maharana, K.; Mondal, S.; Nemade, B. A review: Data pre-processing and data augmentation techniques. *Glob. Transit. Proc.* **2022**, *3*, 91–99. [\[CrossRef\]](#)
21. Nikolopoulos, A.; Karalis, V.D. Implementation of a Generative AI Algorithm for Virtually Increasing the Sample Size of Clinical Studies. *Appl. Sci.* **2024**, *14*, 4570. [\[CrossRef\]](#)
22. Singh Walia, M. Synthetic Data Generation Using Wasserstein Conditional GANs with Gradient Penalty (WCGANS-GP). Master's Thesis, Technological University, Dublin, Ireland, 2020. [\[CrossRef\]](#)
23. Solanki, A.; Naved, M. (Eds.) *GANs for Data Augmentation in Healthcare*; Springer: Cham, Switzerland, 2023. [\[CrossRef\]](#)
24. Bobadilla, J.; Gutiérrez, A. Wasserstein GAN-based architecture to generate collaborative filtering synthetic datasets. *Appl. Intell.* **2024**, *54*, 2472–2490. [\[CrossRef\]](#)
25. Ahmad, Z.; Jaffri, Z.U.A.; Chen, M.; Bao, S. Understanding GANs: Fundamentals, Variants, Training Challenges, Applications, and Open Problems. *Multimed. Tools Appl.* **2024**, *84*, 10347–10423. [\[CrossRef\]](#)
26. Nikolenko, S.I. *Synthetic Data for Deep Learning*; Springer: Cham, Switzerland, 2021; Volume 174. [\[CrossRef\]](#)
27. Cascini, F.; Beccia, F.; Causio, F.A.; Melnyk, A.; Zaino, A.; Ricciardi, W. Scoping review of the current landscape of AI-based applications in clinical trials. *Front. Public Health* **2022**, *10*, 949377. [\[CrossRef\]](#)
28. Karalis, V. Modeling and Simulation in Bioequivalence. In *Modeling in Biopharmaceutics, Pharmacokinetics and Pharmacodynamics*; Interdisciplinary Applied Mathematics; Springer: Cham, Switzerland, 2016; pp. 227–254. [\[CrossRef\]](#)
29. Vlachou, M.; Karalis, V. An in Vitro–in Vivo Simulation Approach for the Prediction of Bioequivalence. *Materials* **2021**, *14*, 555. [\[CrossRef\]](#)
30. Giuffrè, M.; Shung, D.L. Harnessing the power of synthetic data in healthcare: Innovation, application, and privacy. *NPJ Digit. Med.* **2023**, *6*, 186. [\[CrossRef\]](#)
31. Burger, D.A.; Schall, R.; van der Merwe, S. A Robust Method for the Assessment of Average Bioequivalence in the Presence of Outliers and Skewness. *Pharm. Res.* **2021**, *38*, 1697–1709. [\[CrossRef\]](#)
32. Wang, J.; Campbell, G.; Lee, J.H.; Hu, M.; Feng, K.; Chattopadhyay, S.; Zhao, L.; Peck, C.C. Bioequivalence of ANDA Data Using a Non-Informative Bayesian Procedure (BEST) Compared with the Two One-Sided T-Tests (TOST). *AAPS J.* **2025**, *27*, 47. [\[CrossRef\]](#)
33. Ma, H.; Liu, D.; Wu, F. Rectified Wasserstein Generative Adversarial Networks for Perceptual Image Restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 3648–3663. [\[CrossRef\]](#)
34. Yang, L.; Fan, W.; Bouguila, N. Clustering Analysis via Deep Generative Models with Mixture Models. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 340–350. [\[CrossRef\]](#)
35. Lin, Y.-S.; Huang, M.-L.; Li, D.-C.; Yang, J.-Y. Framework Integrating Generative Model with Diffusion Technique to Improve Virtual Sample Generation. *Sens. Mater.* **2024**, *36*, 2439. [\[CrossRef\]](#)
36. Zhang, N.; Ying, S.; Ding, W.; Zhu, K.; Zhu, D. WGNCS: A Robust Hybrid Cross-Version Defect Model via Multi-Objective Optimization and Deep Enhanced Feature Representation. *Inf. Sci.* **2021**, *570*, 545–576. [\[CrossRef\]](#)
37. Lu, L. An Empirical Study of WGAN and WGAN-GP for Enhanced Image Generation. *Appl. Comput. Eng.* **2024**, *83*, 103–109. [\[CrossRef\]](#)
38. Milne, T.; Nachman, A. Wasserstein GANs with Gradient Penalty Compute Congested Transport. *arXiv* **2021**, arXiv:2109.00528. [\[CrossRef\]](#)
39. Papadopoulos, D.; Karali, G.; Karalis, V.D. Bioequivalence Studies of Highly Variable Drugs: An Old Problem Addressed by Artificial Neural Networks. *Appl. Sci.* **2024**, *14*, 5279. [\[CrossRef\]](#)
40. Blow, C.H.; Qian, L.; Gibson, C.; Obiomon, P.; Dong, X. Data Augmentation via Diffusion Model to Enhance AI Fairness. *Front. Artif. Intell.* **2025**, *8*, 1530397. [\[CrossRef\]](#)
41. Ding, B.; Qin, C.; Zhao, R.; Luo, T.; Li, X.; Chen, G.; Xia, W.; Hu, J.; Luu, A.T.; Joty, S. Data Augmentation Using LLMs: Data Perspectives, Learning Paradigms and Challenges. In *Findings of the Association for Computational Linguistics ACL 2024*; Association for Computational Linguistics: Bangkok, Thailand, 2024; pp. 1679–1705. [\[CrossRef\]](#)

-
42. Papadopoulos, D.; Karalis, V.D. Variational Autoencoders for Data Augmentation in Clinical Studies. *Appl. Sci.* **2023**, *13*, 8793. [[CrossRef](#)]
 43. Islam, T.; Hafiz, M.S.; Jim, J.R.; Kabir, M.M.; Mridha, M.F. A Systematic Review of Deep Learning Data Augmentation in Medical Imaging: Recent Advances and Future Research Directions. *Healthc. Anal.* **2024**, *5*, 100340. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.