Scilight

Applied Mathematics and Statistics
https://www.sciltp.com/journals/ams

Article

The Use of Generative AI to Create Hybrid Populations for Bioequivalence Trials

Anastasios Nikolopoulos [1,2] and Vangelis D. Karalis [1,2,*]

[1] Department of Pharmacy, School of Health Sciences, National and Kapodistrian University of Athens, 15784 Athens, Greece
[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

How To Cite: Nikolopoulos, A.; Karalis, V.D. The Use of Generative AI to Create Hybrid Populations for Bioequivalence Trials. *Applied Mathematics and Statistics* **2026**, *3*(1), 3. https://doi.org/10.53941/ams.2026.100003

Received: 23 December 2025
Revised: 27 January 2026
Accepted: 5 March 2026
Published: 27 March 2026

Abstract: Bioequivalence (BE) studies are clinical trials essential in demonstrating therapeutic equivalence across generic and reference medicinal products. Unfortunately, the lack of large samples and broad confidence intervals (CIs) limits their effectiveness. We present a hybrid data augmentation framework in this study that merges actual clinical and synthetic pharmacokinetic (PK) data generation with Wasserstein Generative Adversarial Networks (WGANs). Three randomized, single-dose, 2 × 2 crossover BE datasets (lisinopril, amlodipine, and aceclofenac) were used to develop WGAN-based analyses to directly convert them to virtual subject data. Hybrid datasets were generated by pooling real and synthetic data in fixed proportions while keeping a constant real sample amount. The hybrid datasets were analysed with respect to baseline BE metrics, including geometric mean ratio (GMR), 90% CI, as well as within-subject variability over PK data components, including AUC, Cmax, and the recently proposed average slope (AS). Hybrid datasets achieved a significant reduction in 90% CI widths, with an average reduction of up to 6.7% across all drugs and parameters. For example, in the case of aceclofenac and Cmax, the hybrid 50–150 model reduced the width by approximately 59.8%, decreasing from 18.86% to 7.59%. These results suggest that WGAN-based hybrid datasets can improve the statistical robustness and reproducibility of BE assessments if used as supporting evidence. Whilst clinical data needs to be the foundation for regulatory decisions, hybrid data may be useful for study design minimization and design, sensitivity assessment, and uncertainty reduction particularly in areas where large-scale recruitment is not feasible or an ethically untenable condition.

Keywords: Wasserstein Generative Adversarial Networks; generative AI; data augmentation; hybrid datasets; bioequivalence studies

1. Introduction

An estimation of the sample size required is one of the most common challenges in designing and conducting bioequivalence (BE) studies, and directly impacts the time, cost, and ethical feasibility of the study. Statistical power is critical because BE studies seek to show whether a generic formulation shares the same pharmacokinetic (PK) properties as the drug under consideration. Such underpowered trials could produce wide 90% confidence intervals (CIs) that surpass regulation approval limits (80–125%), leading to inconclusive or failed research [1].

Conversely, when participants should not be overestimated in quantity as they are healthier subjects who, when exposed to the drug of the experiment, do not take medication, leading to high cost, increased ethical issues

at the very least. This is further weakened by the different PKs in the participants, which can mean that sample size requirements are unpredictable [2].

Hence, the design of BE studies often demands careful compromises between practical considerations of ethics and statistical reliability. But this more complex problem has recently been tackled with technology and Artificial Intelligence (AI) breakthroughs, and can lead to more novel solutions. Machine learning algorithms have been applied to a multitude of biological challenges, such as image interpretation and the personalization of medicines and treatments, over the past 10 years [3].

Regarding AI, generative models like Generative Adversarial Networks (GANs), proposed by Goodfellow et al. [4], have exhibited powerful capability in fields including medical imaging, speech synthesis, and electronic health record management [5,6]. Recognizing the features of the data distribution by GANs facilitates the construction of synthetic distributions imitating that of real data, giving a means to augment a sparse set of data. The Wasserstein GANs (WGANs) build upon this work by incorporating the earth mover's distance to its loss function for better training stability and distribution preservation [7,8]. GANs, especially WGANs, may be vital for health care research in countries where there is an acute lack of data.

Despite the remarkable progress of the researchers, conventional methods to calculate the size of study populations from BE analyses mostly use population variability assumptions, which are not necessarily well-served by actual data. This challenge often results in repeated study designs, protocol changes, and duration extensions, which significantly increase costs. Also, more alarming, the need to recruit many healthy volunteers introduces additional ethical challenges. Especially because current clinical data might be more easily used. In this setting, AI-synthetic data augmentation offers a disruptive alternative. It enables scientists to accurately model heterogeneity in the whole population, optimize the design before recruiting, or possibly cut down the number of participants required for any single study.

With today''s latest research and advances in knowledge, however, the state cannot incorporate generated, AI-augmented, or hybrid data as part of BE assessment. The current guidelines of the European Medicines Agency (EMA) [9] and Food and Drug Administration (FDA) [10] state to be developed by the utilization of actual patient data and using existing statistical techniques, including ANOVA and the two one-sided tests procedure. Furthermore, the recently finalized ICH M13A guideline on BE for immediate-release solid oral dosage forms, emphasizes solely on the reliance on actual patient data and the already well-established statistical techniques. It is imperative that AI-generated data be addressed by the regulatory bodies to evaluate their potential benefits and limitations within the BE environment [11]. The old methods and standards, like the previous, are patient-protection measures, while innovative approaches that might work towards efficiency in a clinical trial only exist here. This wide gap between the state-of-the-art technologies available and the actual regulations makes it critical that continued research is conducted to assess whether AI-generated data can be responsibly used to supplement real-life evidence. The rapid AI evolution has moved beyond simple frameworks to include diverse architectures like stable diffusion and transformer-based models, which are increasingly applied to solve complex biological problems. Recent research has demonstrated the utility of these models in generating synthetic medical data, particularly in addressing the limitations of small datasets. These advancements highlight a shift toward more robust data augmentation strategies, setting the stage for their application in highly regulated fields such as BE [12–14].

In this context, previous works from our lab investigated experimental strategies to fill this gap by applying WGANs on BE datasets. We showed that WGAN-generated datasets had statistical characteristics like the original dataset, while the sample one was unable to make the same [15]. The similarity metrics (original versus synthetic, original versus conventional sampling) and statistical power were assessed for the synthetic datasets. This proof-of-concept work showed that AI-based augmentation can make BE evaluations more interpretable and dependable without losing the core of a true clinical study, as they need to be true to the actual clinical data. AI models for data augmentation have been applied in clinical and BE studies on a much larger scale. Historically, WGANs have been used for Monte Carlo simulations such that the generated data remained closely approximating the statistical characteristics of the source population [16], with similar data for previous work. For studies of highly variable drugs, it was shown that the WGAN-based approaches, when compared with reference-scaled EMA and FDA methods, showed equivalent or better performance compared to the regulator-standard. Interestingly, results were improved across larger augmentation ratios, emphasizing the utility of WGAN in highly variable conditions [17]. Likewise, Variational Autoencoders (VAEs) were tested in multiple simulation scenarios, with a consistent increase in numerical power, mainly in analyses that didn''t completely represent the original population attributes. In BE studies, the VAE-generated datasets maintained similar or even better performance compared to the original datasets [18,19]. In the extreme cases of extremely variable drugs, the VAE-generated data again exhibited higher statistical power than the sample datasets, indicating robustness and reliability of the approach [20].

Based on this base, the current study presents a hybrid dataset framework whereby real clinical data are mixed with WGAN-generated data in different percentages for the first time. This approach introduces an innovative technique for using generative AI in business ethics, addressing one of the most enduring challenges in the field: the precise determination of sample size. This objective is addressed by investigating whether and to what extent hybrid datasets will narrow CI widths, better represent the variability of the population, and provide more reproducible BE results by examining pharmacokinetic parameters such as the AUC, Cmax, and a newer average slope (AS) metric [21] (Figure 1). The long-term aim is to provide a theoretical basis addressing not only traditional methodology and statistical constraints on sample size estimation but also to prepare the ground for regulatory negotiations over meaningful integration of AI in the BE trial assessment process. Given the current extent of the need for exploring this type of approach within a BE research context, this study is the first to propose this approach.

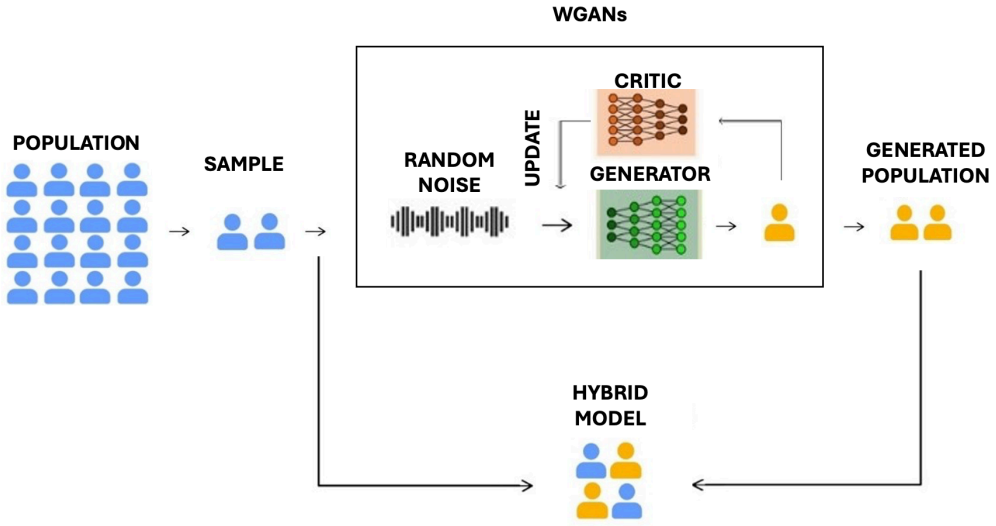

Figure 1. Depiction of the proposed hybrid dataset approach. The sample from the population serves as an input in the WGANs, which then generates a population as an output. The two datasets (sample + generated population) can be joined to create a hybrid dataset, thus enhancing the statistical power and reducing potential errors. A framework illustrating how the power of real clinical data and the power of generative AI may be coupled to produce powerful results.

2. Materials and Methods

The study methodology can be divided into three main components. First, a thorough examination of real BE datasets on which WGANs will be implemented was undertaken. Next is the hyperparameter tuning of the WGAN model. Finally, different types of hybrid models are proposed and assessed that will combine real and WGAN-generated data. By assessing real BE datasets against the limits set by the FDA and the EMA [9,10], the behaviour of the generated data and the hybrid models can be evaluated, compared to the real data, and how well they can co-exist.

2.1. Actual BE Data Used in the Analysis

The first part of the research involved selecting the datasets of the BE studies, served as the foundation of the analysis. Recent studies were deliberately chosen for different types of drugs and number of subjects. The studies share two characteristics: are randomized, single dose, 2 × 2 crossover studies and the subjects were in a fasted condition (Table 1). The number of subjects (N) represents those who completed the BE study after initial screening, application of exclusion criteria, and any protocol violations or partial withdrawals.

Table 1. Summary of the BE studies used as input data for WGANs.

BE Study	Li et al. [22]	Li et al. [22]	Bushra et al. [23]
Study Type	Randomized, single dose, 2 × 2 crossover study	Randomized, single dose, 2 × 2 crossover study	Randomized, single dose, 2 × 2 crossover study
Drug Name	Lisinopril	Amlodipine	Aceclofenac
Drug Type	ACEI	Dihydropyridine-based CCB	NSAID
Fed/Fasted Conditions	Fasted	Fast	Fasted
N (Completed)	39	39	12

Key: ACE, angiotensin-converting enzyme inhibitors; CCB, calcium channel blockers; NSAID, non-steroidal anti-inflammatory drugs.

The research is focused on the following pharmacokinetic parameters: area under the curve (AUC) and peak drug concentration (Cmax), which are depicted in the following figure (Figure 2a). Additionally, the AS was calculated for each dataset, based on recent research, where AS exhibited the necessary properties, including kinetic sensitivity and proper units, making it a more reliable metric compared to the Cmax, which primarily reflects the extent rather than the rate of absorption and lacks appropriate kinetic units (Figure 2b) [21,24].

The 90% CI width was calculated as the difference between the upper and lower confidence limits of the 90% CI for the geometric mean ratio of the ln-transformed AUC and Cmax, as specified in the FDA and EMA guidelines on the Investigation of BE [9,10].

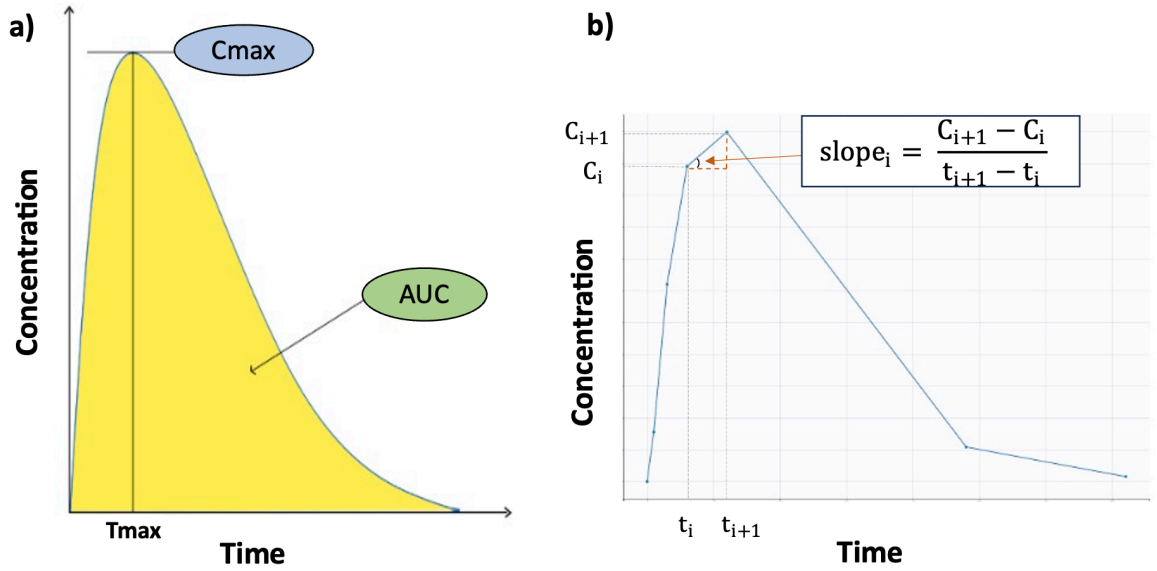

Figure 2. Visual representation of the parameters examined in this research: (**a**) Area under the curve (AUC), which is the integral of the plasma concentration-time curve, and highest concentration (Cmax), which refers to the peak level the drug achieves in the body. Both are derived from the plasma concentration-time curve, as depicted. This curve is generated by giving a drug to a subject, taking blood samples at various time points, and measuring the amount of tested drug present in the blood plasma; (**b**) AS, which is calculated by the average of the tangents between two consecutive points [21].

2.2. WGANs and Hyperparameter Tuning

GANs consist of two neural networks, a Generator (*G*) and a Discriminator (*D*), that are trained against each other to improve the quality of data generation, by following a two-player minimax game with a value function $V(D,G)$ (Equation (1)) [4,25]:

$$\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}\left[\log D_{(x)}\right] + E_{z\sim p_z(z)}\left[\log\left(1 - D\left(G_{(z)}\right)\right)\right]. \quad (1)$$

G's purpose is to create, from random noise ($p_z(z)$), convincing fake data that looks like real data ($p_{\text{data}}(x)$), by learning the underlying distribution of the real data and mimicking it. On the other hand, *D*'s goal is to correctly distinguish the real from the fake data that were produced by the *G*:

(1) $D_{(x)} \approx 1$, meaning that *D* thinks the sample is real, or
(2) $D_{(x)} \approx 0$, meaning that *D* thinks the sample is fake

During this minimax game, *G* and the *D* compete against each other until they reach an equilibrium. *D* must correctly classify both real and fake samples, while *G* wants to minimize the performance of the *D* on the fake data.

However, GANs face the problem of unstable training, which can be a major hurdle and can result in mode collapse, in which the generator produces similar outputs. In order to address this, WGANs introduce the Wasserstein distance (Equation (2)) [7] as their loss function, offering smoother gradients that increase training stability and convergence [25]:

$$W\left(\mathbb{P}_r, \mathbb{P}_g\right) = inf_{\gamma\in\Pi\left(\mathbb{P}_r,\mathbb{P}_g\right)}\mathbb{E}_{(x,y)\sim\gamma}[\|x-y\|], \quad (2)$$

where:

- $\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)$ means that γ serves as a joint probability distribution that specifies how much probability mass should be moved from every point x in $\mathbb{P}_r$, to every point y in $\mathbb{P}_g$.
- The constraint $\Pi(\mathbb{P}_r, \mathbb{P}_g)$ ensures that with the sum of all the mass leaving from $\mathbb{P}_r$ you get $\mathbb{P}_r$ itself, and with the sum of all the mass arriving at $\mathbb{P}_g$ you get $\mathbb{P}_g$ itself (conservation of mass).
- $\|x - y\|$ refers to the distance (typically Euclidean) between the two points x (source) and y (destination). The cost is the distance the mass must travel.
- $\mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|]$ refers to the calculation of the expected total cost for a given transport plan γ.
- The final term $inf_{\gamma\in\Pi(\mathbb{P}_r,\mathbb{P}_g)}$ stands for the infimum, the greatest lower bound, so that the equation should iterate through all possible transport plans and then select the one that results in the minimum total cost ($W(\mathbb{P}_r, \mathbb{P}_g)$).

This loss function helps to determine how different the generated data are from the real ones [26,27]. This method keeps the discriminator from becoming too strong, resulting in balanced training between the two networks. The use of Wasserstein distance in WGANs helps in a better alignment of generated and real data distributions, which improves the quality of synthetic data [26]. Concluding, WGANs showcase a more robust framework in generative AI, which makes them well-suited for applications that require stable training and high-quality data generation.

The architecture of the WGAN model and the hyperparameter tuning are of utmost importance, because they can directly impact the ability of the generated data to co-exist with the real data. In the following table (Table 2) the different values that were evaluated for each hyperparameter or parts of the WGAN architecture. After thoroughly examining different scenarios, the optimal combination for the model was selected. Finally, to ensure consistency, robustness, and reproducibility, these parameters remained constant throughout the whole research process.

Table 2. Architecture of the WGAN algorithm and the different values that were examined during the hyperparameter tuning of the model.

Generator-Critic	Hidden Layers	1		
	Neurons in Hidden Layers	4	6	8
Activation Function	Hidden Layers	linear	ReLU	
	Output Layers	sigmoid	tanh	linear
Learning Rates		0.01–0.00001		
Latent Dimensions		80	100	120
Batch Size		32	64	128
Epochs		150	300	500

Based on the hyperparameter tuning summarized in Table 2, the final WGAN configuration used one hidden layer with 6 neurons in both the generator and the critic. A latent space dimensionality of 100 was used, with ReLU activation functions in the hidden layers and linear activation at the output layer to enable unconstrained generation of continuous pharmacokinetic variables. A learning rate of 1×10^{-4}, batch size of 64, and 300 training epochs were used for model training. In order to provide reproducible and stable generative processes, such as in this paper, the WGAN architecture, as well as any hyperparameters attached to it, was kept the same across all drugs and pharmacokinetic parameters studied. A fixed random seed was used for the initialization of the network, updates of weights, and shuffling data so as to minimize random initialization-induced stochastic variability. Each training and generation procedure was repeated multiple times under exactly the same conditions. They were then used to examine the convergence stability of outputs obtained from the generator itself, and all those produced are then tested in general. Reproducibility and stability were controlled similarly, as well, in our previous work, where the following repeated Monte Carlo simulations, extensive hyperparameter tuning, and distributional similarity evaluation between original, resampled, and generated data were applied and validated [15–17]. These iterative methods reduce the effect of random variability inherent in GAN training, help in the detection of convergence instabilities of the model or mode collapse, and confirm the reported hybrid data features are representative of stable model behavior, independent of each random fluctuation that occurs in the data.

In addition, we should note that the WGAN model was independently trained on each investigated drug (lisinopril, amlodipine, and aceclofenac) separately. For all studies, the respective AUC, Cmax, and AS pharmacokinetic summary parameters (as indicated in original BE data sets [22,23]) were included in the inputs for the model. Its raw concentration–time profiles were not used for processing, nor was it used to model physiological pharmacokinetic constants (bioavailability or absorption rate constants). It concentrated solely on the direct generation of the BE assessment parameters used for the following statistical analyses.

2.3. Hybrid Model Approach

This study examines hybrid models that combine real sample data with WGAN-generated distributions ("virtual patients") in varying proportions. Their performance was evaluated by holding a fixed real sample size and increasing the overall hybrid dataset size as the generated data was added. Figure 3 illustrates the technique: a random selection of half of the available dataset (i.e., half of the subjects from the original BE dataset) was used as input to the WGAN model. The model then created a synthetic population of "virtual patients", which was combined with the sample to form hybrid models where real and generated data coexist.

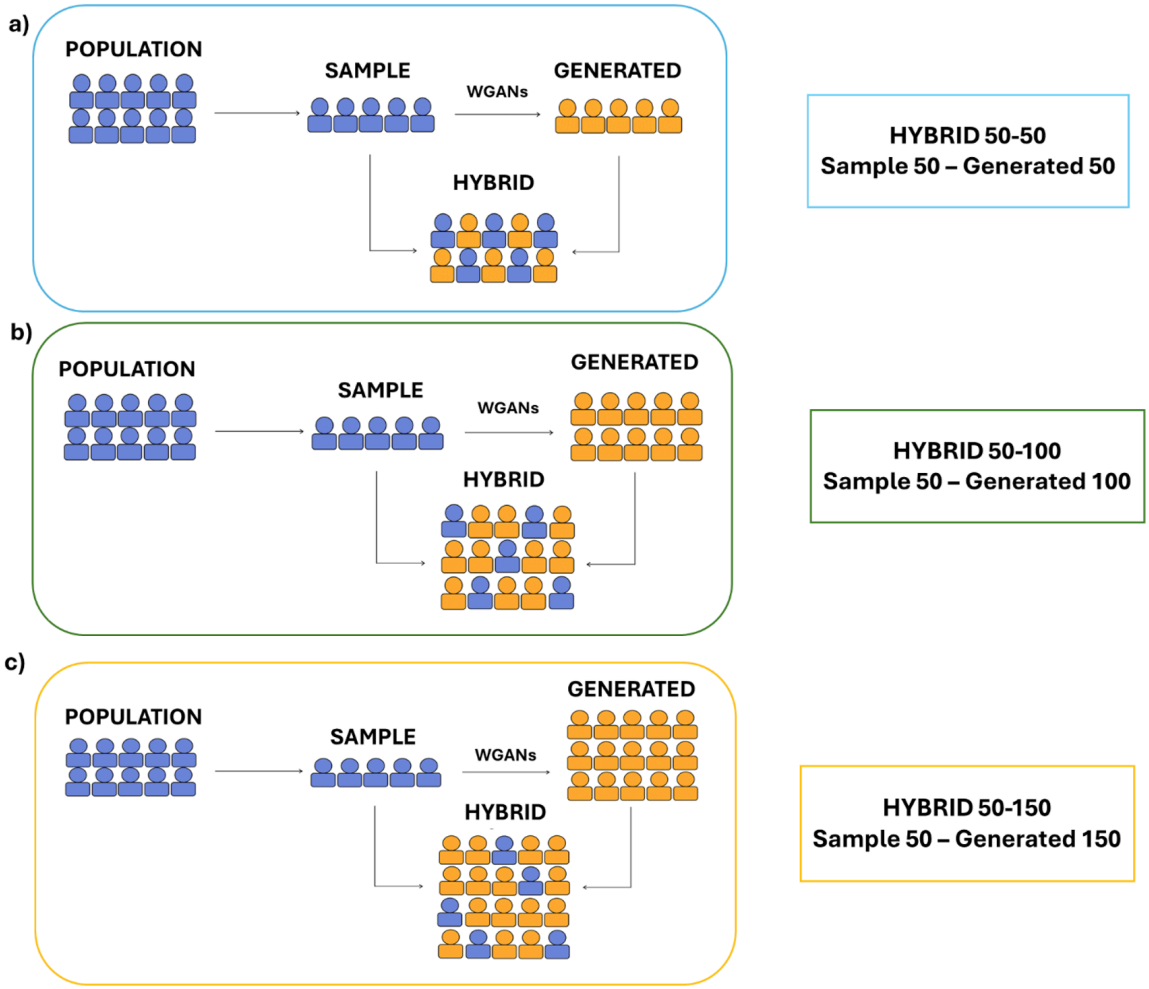

Figure 3. Illustration of the hybrid dataset approach introduced in this study: (**a**) the hybrid 50–50 approach combines equal proportions of sample and generated data; (**b**) the hybrid 50–100 approach includes twice as much generated data as the sample dataset; and (**c**) the hybrid 50–150 approach incorporates three times the generated data compared to the sample dataset. The three different ratios serve as a first step for assessing the feasibility of such a methodology and showcasing the performance of using a greater amount of generated data in each hybrid model.

An examination of three types of hybrid models was undertaken. The "Hybrid 50–50" consists of an equal mix of real and generated data. The "Hybrid 50–100" and "Hybrid 50–150" expand the dataset by two and three times the sample size, respectively. These variations allow us to assess the extent to which real and generated data can coexist effectively and how their interaction changes as the proportion of generated data increases. For each hybrid scenario, the metrics of the corresponding generated one were also noted to provide a total perspective of the performance of the WGANs.

For the evaluation of each distribution (original, sample, generated, and hybrid), the metrics calculated included the geometric mean ratio (GMR), 90% CI (upper and lower limits, as well as width), and within-subject coefficient of variation (CVw), which was calculated through the residual variability of the ANOVA analysis. These metrics were selected due to their critical role in assessing the BE of a study.

3. Results

As previously stated, certain metrics were calculated for each drug and pharmacokinetic parameter in order to evaluate the distributions' performance. The findings for each parameter were collected in separate tables, and the corresponding figures serve as a visual depiction of the 90% CI widths between the original and hybrid distributions. Figure A1 shows histogram-based comparative observations of original, sparsely sampled, and WGAN-driven datasets, presenting a controlled demonstration of distributional similarity and generative fidelity amidst differing Test–Reference discrepancies.

Table A1 contains the AUC parameter results. For lisinopril, the GMR and CVw values for all hybrid datasets were closer to the originals, whereas the sample dataset's values were greater than both the original and hybrid values. Amlodipine behaves similarly, with the GMR and CVw corresponding more closely to the original than the sample dataset. For aceclofenac and the GMR metric, only the hybrid 50–100 scenario was closer to the original. However, the CVw for all hybrid datasets were more like the original dataset than the sample to the original dataset. In total, for most cases, the hybrid scenarios achieved metrics much more related to the original dataset than what the sample could achieve. Figure 4 illustrates the comparison of the 90% CI widths. Amlodipine and aceclofenac perform similarly in the hybrid 50–150 scenario, having the smallest width compared to the other scenarios and the original. On the other hand, the widths of lisinopril are narrowing, except for the hybrid 50–150, which is slightly broader than the hybrid 50–100 scenario.

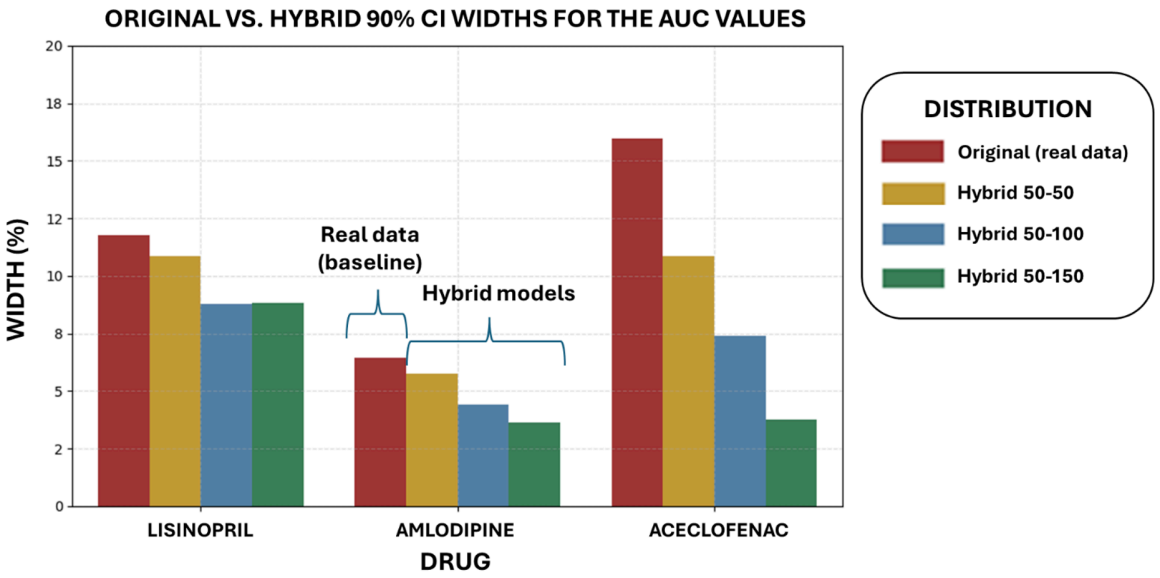

Figure 4. Comparison of the 90% CI widths for the AUC values between original and hybrid distributions for each drug. Here, the original distribution (real data) serves as a baseline next to the results of the hybrid models. For each type of drug, real data showcase a bigger width percentage as opposed to the hybrid models. The figure also demonstrates the reduction in width percentage in most case scenarios (except for the hybrid 50–150 of lisinopril).

Table A2 shows the Cmax parameter results. To begin, lisinopril behaves similarly to the AUC metric, with the GMR and CVw values for all hybrid datasets closely resembling the original dataset rather than the sample. Except for the hybrid 50–100 scenario, the GMR values for amlodipine are identical to those for the original. In terms of CVw, however, all hybrid scenarios resemble the original more than the sample does. Finally, for aceclofenac, the GMR and CVw metrics in the hybrid situations are more closely connected to the original values than the sample is. Once again, most hybrid scenarios could better capture the original dataset's characteristics than the sample dataset's ones.

Figure 5 provides the same information as Figure 4, but for the Cmax parameter. It is clear that all hybrid situations, regardless of drug type, had a narrower width than the original dataset. As the percentage of generated data increases, the width decreases.

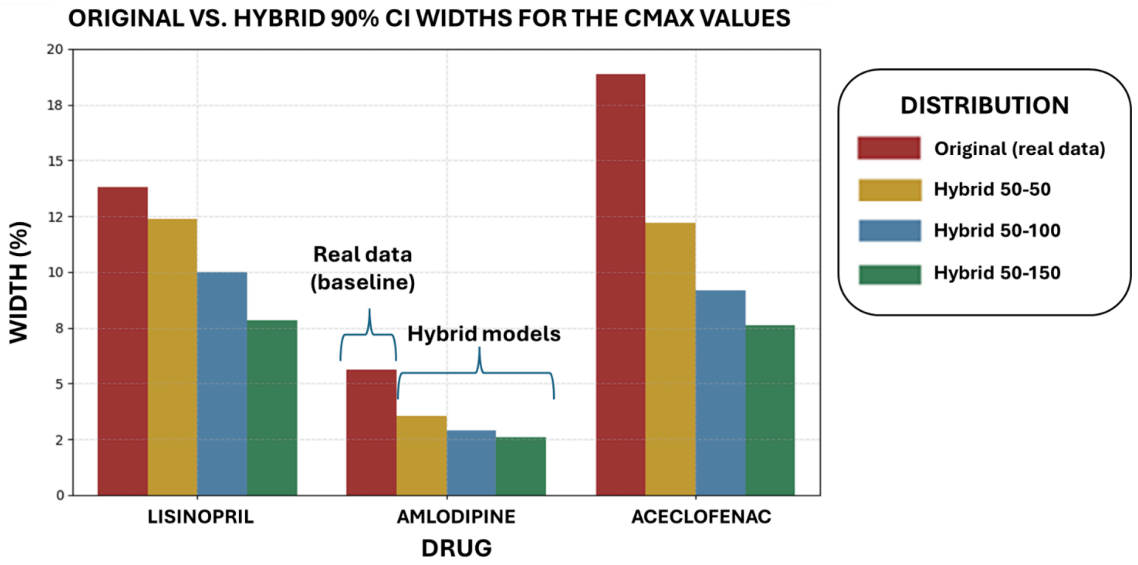

Figure 5. Comparison of the 90% CI widths for the Cmax values between original and hybrid distributions for each drug. Here, the original distribution (real data) serves as a baseline next to the results of the hybrid models. For each type of drug, real data showcase a bigger width percentage as opposed to the hybrid models. The figure also demonstrates the reduction in width percentage as the hybrid model incorporates more generated data.

Table A3 shows the results for the AS parameter. Starting with lisinopril, the hybrid scenarios for both CVw and GMR metrics appear to more closely mirror the original dataset than the sample. GMR and CVw values are closer to the original in two of the three hybrid scenarios for amlodipine (hybrid 50–50 and hybrid 50–150), whereas in the hybrid 50–100 scenario, the CVw is closer to the original, but the GMR is not. Finally, for aceclofenac, all hybrid scenarios have GMR values closer to the original dataset; however, only the scenario hybrid 50–150 has a CVw value closer to the original than the sample does. To summarize, most hybrid scenarios' values were closer to the original dataset than the sample's.

Figure 6 follows the logic of Figures 4 and 5, but with the AS parameter. The widths of all hybrid scenarios were smaller than the originals, with the width decreasing as the hybrid scenario contained more generated data, for all types of drugs.

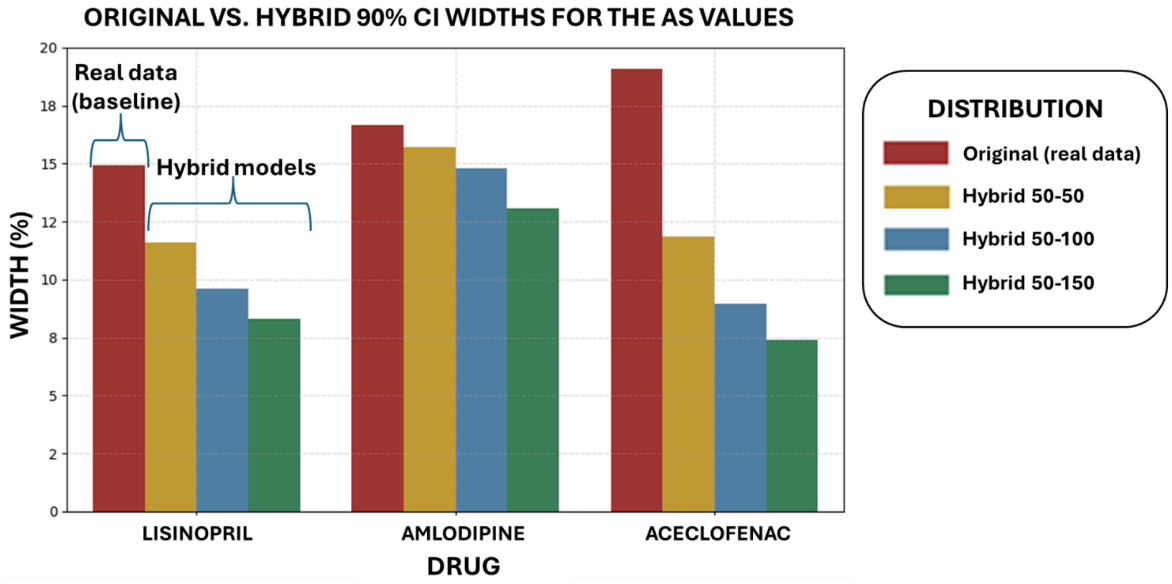

Figure 6. Comparison of the 90% CI widths for the AS values between original and hybrid distributions for each drug. Here, the original distribution (real data) serves as a baseline next to the results of the hybrid models. For each type of drug, real data showcase a bigger width percentage as opposed to the hybrid models. The figure also demonstrates the reduction in width percentage as the hybrid model incorporates more generated data.

Figure 7 depicts the average performance of each hybrid scenario in terms of 90% CI widths compared to the original ones, for all types of drugs and different parameters. This figure serves as an efficiency metric to depict the behaviour of the different hybrid scenarios. The average width difference for the 50–150 hybrid scenario is clearly the largest, and it increases as the percentage of generated data in the hybrid dataset increases.

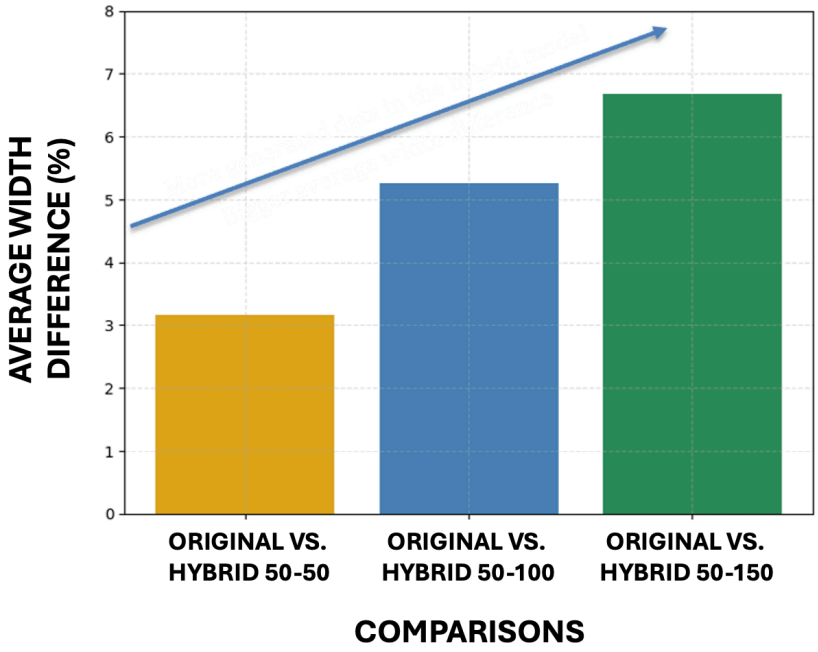

Figure 7. Comparison of the hybrid models against the original distributions. The widths of 90% CI were taken into consideration for all different types of drugs (lisinopril, amlodipine, aceclofenac) and the three parameters examined in this study (AUC, Cmax, AS). For each pair of calculations (e.g., lisinopril-AUC-original width minus lisinopril-AUC-hybrid 50–50 width), the values were noted and later the average difference for each one of the three different hybrid models (50–50, 50–100, 50–150) was calculated.

4. Discussion

WGANs were shown to have significant advantages for BE datasets by existing studies, where they can effectively produce synthetic pharmacokinetic profiles that are comparable to the original population [15]. Synthetic data in the study reproduced essential population characteristics while narrowing the size of the 90% CIs for that group of patients, showing that AI-based data augmentation may be a promising complement to traditional BE analyses. But those scenarios were primarily concerned with fully generated datasets and did not explicitly outline the drawbacks of synthetic data alone, or the potential benefits of integrating generated data with actual clinical observations. Here, we build on this methodological backbone with hybrid data sets that combine real clinical data with WGAN-generated data in three fixed proportions. We tested this strategy with three randomized, single-dose, 2 × 2 crossover BE studies involving lisinopril, amlodipine, and aceclofenac [22,23]. AUC, Cmax, and the recently proposed AS [21,24], which are key PK parameters, were investigated. The basic question was that if hybrid data sets could achieve greater reductions in CI widths and greater reproducibility, population-level statistical characteristics should be more than just one-hot subsampling/generation methods allowed.

Furthermore, under the current rules of the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA) that have been promulgated recently, BE findings should reflect real clinical data, analyzed using established statistical methodologies [9,10]. Moreover, while the ICH M13A guideline aims to harmonize BE study designs, it primarily focuses on traditional clinical frameworks, without stating anything regarding the use of synthetic data generated from AI models [11]. It is this point that stands out most. Despite the accelerated development of generative models in biomedical research, regulatory science has not yet delineated formal mechanisms to include AI-derived data in a confirmatory clinical assessment. However, there is increasing acknowledgement of the value proposition for the role artificial intelligence in drug development and clinical trial design has for this type of evidence, use of novel compounds [28,29], evidenced through pilot programs and reflective papers that focus on the ethical considerations and methodology relevant to the fusion of machine learning and empirical data [30,31].

On this changing horizon, this paper provides a practical use case that illustrates how AI-supported data, when openly integrated with real datasets, can improve the statistical robustness of BE analyses, while respecting the supremacy of clinical outcomes. An overarching principle underlying this effort is that AI-generated data should complement rather than replace actual clinical data. The findings indicate that while generated-only datasets frequently mimic the original population in summary statistics, they are not clinically proven and are not suited for regulatory acceptability, so they cannot contribute to BE conclusions independently. In contrast, hybrid datasets exhibited narrower 90% CI widths per drug and PK parameter investigated, but with central tendency estimates on the same level as the original data.

In the three drugs involved, hybrid augmentation was consistently beneficial. When comparing both lisinopril and amlodipine, hybrid data sets resulted in a better GMR and CVw accuracy, typically more similar to the original populations than the sample-based analyses. For aceclofenac, we found that the hybrid approaches were especially effective at CIs contraction for Cmax and AS, and thus they demonstrate the capacity to accommodate heterogeneity in compounds that may display different absorption and variability profiles. Despite the parameter and drug dependence of alignment with population variability, the general tendency provides evidence of the generalization effect of hybrid augmentation from several PK properties.

Figure A1 presents histogram-level comparisons based on simulated and not the clinical models to validate the inherent generative power of the WGAN framework under fully controlled conditions by using simulated rather than the original clinical datasets. As we employ simulated populations, the mechanism by which true data is generated–such as population distributions, effect sizes, and variation structures–can be clearly expressed a priori, an impossible condition for real clinical datasets. This allows for a definitive assessment of a WGAN ability to retrieve known underlying distributions and not only to provide typical idiosyncratic features from proprietary trial data. Two established Test–Reference ratios (a = 1.00 and a = 1.05) are also included, ensuring a direct assessment of model behaviour under both bioequivalent and marginally non-equivalent conditions and demonstrating how differences in Test and Reference formulations are propagated by sparse sampling and generative augmentation. This controlled experimental environment for distinguishing fidelity from CI narrowing, however, is very important as it makes clear that the observed stabilization of pharmacokinetic metrics, which is a typical effect of generalisation, does not represent an artifact of data enrichment, but instead reflects true reconstruction of underlying distributions.

One fundamental methodological issue is that just increasing the data size will automatically produce narrower CIs, so you cannot use the data to show the specific benefits of the WGAN enhanced feature on its own. To deal with this issue, our prior research extensively compared WGAN-generated data sets to conventionally resampled reference datasets matched for sample size, revealing that WGANs outperform naïve resampling with respect to preserving population-level features like geometric mean ratios and CVw, especially in small sample and highly variable conditions [15–17]. These studies also carried out highly elaborate hyperparameter tuning, repeated Monte Carlo simulations, and fixed random seeds to evaluate convergence stability and reproducibility for the independent runs. Expanding on this well-established framework, the current study follows identical architecture and reproducibility guarantees while adding hybrid real–synthetic datasets from an authentic BE experiment to the analysis. As shown in Figure 8, modest disparities in point estimates (especially for the AS) suggest that in short-run training data, the error structure of AS is more sensitive, local, and to the residual error than the generative model causes due to systematic bias [24]. In other words, this deviation affects point-estimate stability rather than BE decision boundaries and therefore does not imply inflation of type I error. This behaviour for this approach will allow a clearer, more systematic testing of the method by repeating experiments over several random seeds and by augmentation ratios, and will allow a more complete assessment of the effect, distilling stochastic sampling from actual generative performance.

Tables A1–A3 also demonstrate that generated-only datasets offer rich contextual information but cannot supplant the clinical basis necessary for BE work. Hybrid models, in contrast, bring together the strengths of both worlds: the truthfulness of real clinical data and the statistical power of AI-enhanced augmentation. These results can have critical implications in terms of sample size. Although all analyses have been conducted with a fixed real sample size, the results indicate that a significant subset of actual subjects might keep important population characteristics for exploratory and design-stage models and analyses when used on the basis of generative data at the design and exploration level. In Li et al. [22], 39 subjects completed the trial, of whom 36 were used for model training; exploratory analyses suggest that a subset of 18 real subjects, when complemented with virtual patients, may preserve important population characteristics at the design and sensitivity-analysis stage, without implying any reduction in sample size requirements for confirmatory regulatory BE evaluations. Similarly, in the aceclofenac study conducted by Bushra et al. [23], hybrid models similarly behaved from a subset of 6 subjects.

These observations pertain to study planning and sensitivity analyses and are not indicative of a scaling down of sample size needed for confirmatory regulatory evaluations.

To make these new findings relevant for the field of regulations, the guidelines that support the use of generative models in BE need to be transparent. On a technical level, convergence diagnostics, activation functions, learning rates, and batch sizes for GAN design parameters are to be standardized and uniformized for reproducible and homogenous results [32–35]. Challenges from a research standpoint could be the ratio of synthetic data to the real (for reporting, comparisons to independent validation data), as well as a number of hybrids (e.g. 50–50, 50–100) for model robustness [36–38]. Model architecture transparency, random seed control, and independent replication of findings are also critical [39,40]. Such steps would ultimately contribute to establishing that the hybrid models are scientifically sound and practical.

Despite the potential benefits of the presented hybrid approach, several significant limitations should be identified. First, the current study included only 3 BE datasets based on randomized, single-dose, 2 × 2 crossover studies conducted in fasting conditions, thereby limiting the generality of the findings. Prior, however, to generalize these findings more broadly, it is necessary to validate these claims on other drug classes, formulations, dosing schedules, fed conditions, and study designs. Second, the WGAN generalization performance is sensitive to architectural decisions, hyperparameter adjustment, and random initialization: poor training or unstable convergence can introduce subtle bias, mode collapse, or variances that summary statistics alone wouldn't reveal. The third problem concerns the generation of pharmacokinetic summary metrics (AUC, Cmax, and AS) rather than full concentration–time profiles, which restricts mechanistic interpretability and may obfuscate clinically relevant temporal dynamics of absorption and elimination. Fourth, the AI-assisted data set for regulatory application contains ethical/legal issues related to transparency, accountability, and reproducibility (e.g. [41–43]), emphasizing the importance of traceable model development pipelines and uniform reporting. A further important limitation related to statistical power may be considered in raising statistical power and the potential of false positive BE results. Hybrid augmentation diminishes CIs but is not to be exploited to circumvent stringent sample-size restrictions. Conversely, generative AI should be treated like a sensitivity analysis and study-design optimization tool. Similarities and replicability of original distributions, generated distributions, and sampled distributions have been found to reduce false-positive rates and help maximize statistical robustness [15,16], and the inclusion of these safeguards should be systematically incorporated in future implementations. Additionally, it is important to distinguish whether CI narrowing is a real reappearance of variances at the population level or just the mathematical result of expanded dataset size. The current study shows that hybrid datasets preserve geometric mean ratios and CVw nearer to those of the original populations as compared to small empirical samples alone, but independent datasets and failed BE studies are necessary for further validation. Lastly, many regulatory bodies already take action against high pharmacokinetic variability with scaled acceptance limits, and WGANs have previously been examined in simulated extreme-variability scenarios [17]. The hybrid approach does not intend to mitigate abysmal formulation performance but to minimize uncertainty in the more borderline locations, in which recruitment is either ethically deficient or logistically difficult, to position it as a complementary, not a substitutive method within the existing BE paradigm.

The ethical implications of blending AI-generated synthetic data into BE studies, particularly one focused on the partial replacement of actual human subjects, must be balanced effectively by the imperative to maintain evidence of clinical quality. So, in the present framework, synthetic data are introduced at the design and sensitivity-analysis stage to complement a defined proportion of enrolled subjects, while all confirmatory inferences remain exclusively based on empirical clinical data. The goal is that the use would reduce uncertainties and improve our statistical efficiency, while still keeping the clinical heart of this research firmly in patients' eyes. All confirmatory inferences are still based on the empirical data in accordance with EMA, FDA, and ICH M13A, and hybrid datasets are employed as auxiliary tools for sensitivity analysis, uncertainty quantification, and design optimization. By keeping a strict separation between AI-driven augmentation and regulatory evidence generation, the approach attempts to avoid unnecessary exposure of healthy volunteers without upsetting either ethical standards or the scientific standards of valid BE results.

Apart from the methodological implications, there is significant applicability of hybrid augmentation. Restricting the enrolment might cut study costs, logistical challenges, and may prevent unwanted risk to healthy volunteers. It is shown in Table 3 that using a fictitious WGAN-assisted design for small-molecule drugs, both pilot and pivotal studies, compared the traditional two-stage BE design [44]. The reduced number of participants and cost per subject outlined in the table were informed by simplified assumptions meant to help contextualize the potential operational significance of hybrid modalities, not empirical (validated) estimates. However, such data are illustrative, not representative of the regulatory acceptability of a potential loss in enrolling participants with

BE in confirmatory BE trials. Nonetheless, they also provide one useful theoretical model of what might change in practice because of hybridization in terms of trial efficiency.

Table 3. Illustrative comparison between a typical BE study and a hybrid (50–50) scenario. Shown are example sample sizes and estimated total study costs for a conventional BE design using only real subjects and for a hybrid design in which 50% of subjects are supplemented by WGAN-generated virtual subjects. Cost estimates are based on a commercial analytical context, at a cost per sample of $200 (within the price range of $100–350 per sample), 12 blood samples per period, and two dosing periods in a standard 2 × 2 crossover design. The "Estimated cost savings" column captures the difference between the estimated costs in both cases and the anticipated cost savings of lower recruitment.

Typical Study		Hybrid Study (50–50)	
Sample Size	**Cost ($)**	**Hybrid Study**	**Estimated Cost Savings ($)**
24	115,200	12	57,600
36	172,800	18	86,400
48	230,400	24	115,200
72	345,600	36	172,800

The conceptual benefit of hybrid datasets is depicted in Figure 8, where 90% CI widths for lisinopril are assessed across real, sample-based, and hybrid data. Sample-based analyses lead to larger CIs, resulting in more potential for inconclusive or borderline findings in genuine BE cases. Hybrid datasets provide narrower CIs, which fall within regulatory acceptance thresholds, but still exhibit the central behavior of the data. This visualization highlights how hybrid augmentation may reduce the uncertainty in the decision-making process and increase robustness of the decisions, particularly for difficult or borderline BE cases, without compromising regulatory principles.

The contributions of the study rely on three major pillars. The first one is the introduction of a hybrid framework for the first time, which merges real clinical data with AI-synthesized virtual patients. The second one is the comparison of the behaviour of the sample to that of the hybrid models, where samples failed consistently to capture the characteristics of the original dataset. Third, the newly introduced pharmacokinetic parameter AS is used as a BE assessment metric to prove its utility beyond traditional PK parameters. Lastly, the estimated cost benefits after applying the proposed AI approach to reduce sample sizes are presented.

Further studies may widen the applied hybrid framework with different drug classes, formulations, and study designs to establish that hybrid-based model generalizability and operational robustness are feasible. In particular, physiologically based pharmacokinetic (PBPK) modelling could offer mechanistic interpretability to the generated hybrid populations [45], and conditional and stratified generative architectures could be developed and implemented to simulate clinically relevant subgroups [46]. In addition, pilot partnerships with regulatory agencies are required to co-develop validation benchmarks, governance principles, and transparency standards for responsible AI-augmented data as part of BE research [47]. Importantly, future validation efforts should include real-world failed BE studies to confirm that the WGAN–hybrid framework properly identifies non-bioequivalent formulations and does not inadvertently inflate false-positive rates, protecting the scientific and regulatory credibility of the approach.

In the same vein, there are a number of other experimental tasks that are suggested to validate the conclusions of the current study. For one, extended external validation over more BE datasets of multiple drug classes and variability profiles is necessary for determining the generalizability of the hybrid augmentation approach beyond the three examples that are presented here. Second, negative controls with real or simulated studies, both of which have been reported and demonstrated to fail BE, should be included to explicitly consider the risk of false positive findings and control for type I error. Third, we suggest testing robustness, sensitivity, and analyses with respect to augmentation ratios, random initialization, and WGAN hyperparameters to estimate the stability of outputs and the effect of model assumptions on findings. Fourthly, a comparison to competing augmentation baselines (e.g., traditional resampling or variational autoencoder-based generation) would serve as a useful benchmark to evaluate the relevant merits of the WGAN-based method. Last but not least, it would be interesting for the work to augment summary BE statistics with quantitative distributional similarity measures between real and synthetic data (e.g., distance-based and tail-sensitive diagnostics) to guarantee that CI narrowing is a faithful reconstruction of population variability, and not simply a mathematical artifact of expanding the number of samples provided.

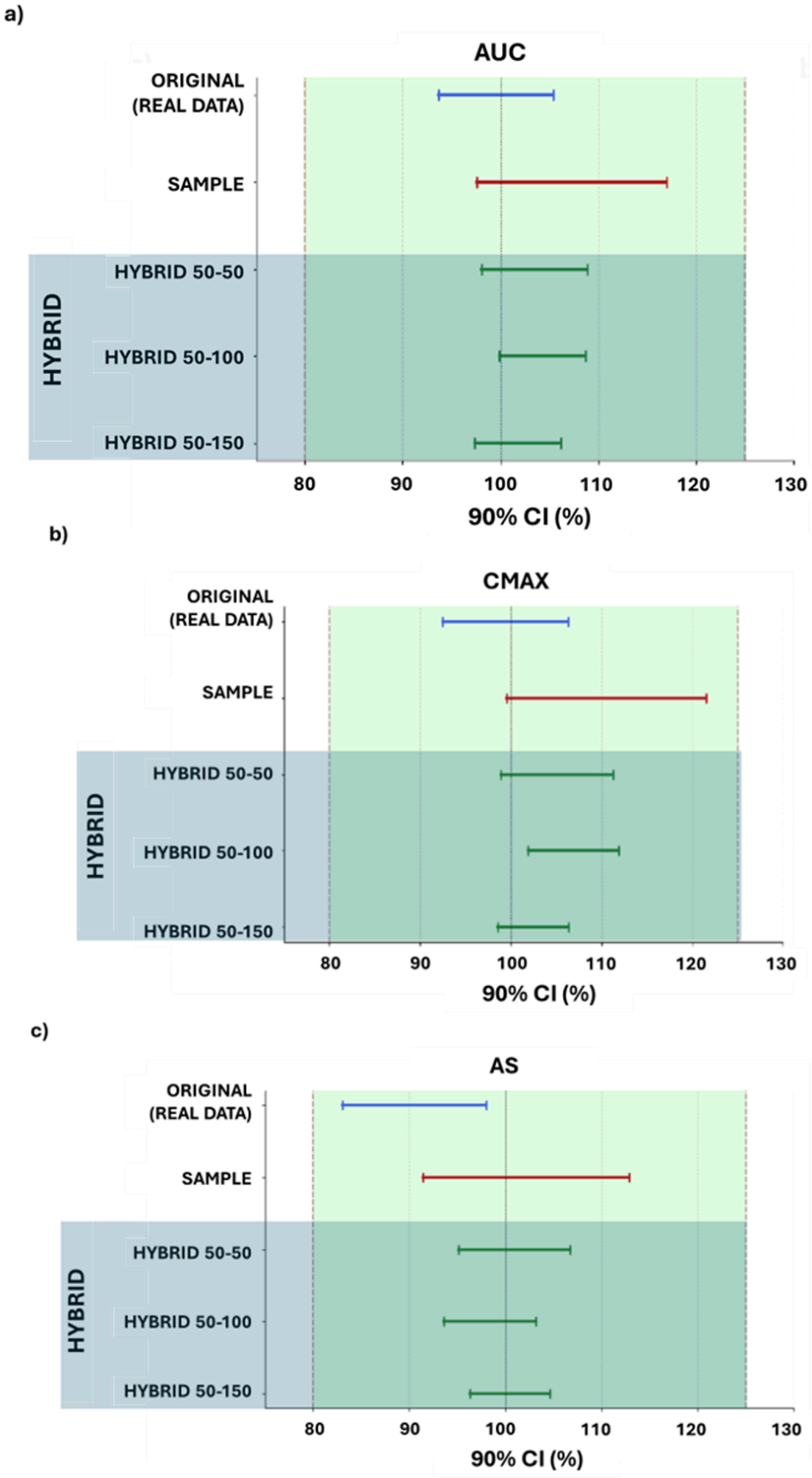

Figure 8. A graphic representation of the proposed approach discussed and the combined results for each pharmacokinetic parameter for lisinopril. The green field represents the acceptable limits (80–125%) in the analysis of a BE study. This approach proposes that a hybrid dataset, consisting of real clinical and AI-generated data, can provide a much narrower range of the 90% CI, still representing the characteristics of the real data. With blue, the widths from the real data are represented, with red the ones from the samples (highlighting the fact that they are wider), and with green the ones of the hybrid models.

From a regulatory and public-health perspective, the key goal of BE assessment should be the strict control of type I error, the risk of making misinformed BE decisions on non-bioequivalent formulations. While we found that hybrid augmentation stabilized estimation of parameters and diminished the CI widths, some open methodological and regulatory issues inhibit a feasible implementation of such an approach. In BE studies, large sample sizes are often necessary under two main conditions: (a) drugs display high CVw, evident in the large residual error terms in the ANOVA model, as in highly variable drugs; (b) the expected geometric mean ratio is far removed from unity. To address excessive sample size requirements for highly variable drugs, regulatory agencies have presented reference-scaled average BE methods. Yet these approaches require replicating clinical study designs, bringing multiple statistical complexities into play, and further increasing drug exposures per volunteer even in the presence of a smaller total number of participants. Thus, while the subject numbers can be decreased, the individual burden on each participant increases. On the other hand, AI-assisted hybrid augmentation is more conceptually easier, and therefore easily reproducible, to stabilize estimates of variability and eliminate uncertainty from the design phase while also not imposing an additional burden on the participants in clinical practice. However, these benefits must not undermine consumer protection. Future implementations of the WGAN–hybrid framework should therefore systematically integrate negative controls, such as failed real-world BE studies and simulated non-bioequivalent scenarios, to directly assess the framework's effect on false-positive rates. Such appraisals are necessary to establish that observed decreases in CI widths represent better modelling of population variability, as opposed to a statistical artifact arising from data expansion and the preservation of the core regulatory aim of the control of type I error.

Collectively, these findings confirm hybrid datasets as a feasible and scientifically validated approach to improving the statistical integrity and replicability of BE studies. While generated data alone is not sufficient for regulatory decision-making, the hybrid models achieve a compromise between clinical evidence and statistical efficiency. In a real world BE study with a small sample, the mean often shifts due to the 'luck of the draw' in participant recruitment. Figure 8 illustrates that while a small sample might lead to a borderline BE success result, the hybrid model provides a more robust estimate that accounts for population-level variability, resulting in the reduction of false-negative results. To use an analogy, the hybrid model can serve as a stabilizer that incorporates WGAN generated data that reflect the broader population distribution, which results in CI that are not only narrower but also more centered toward the true population characteristics compared to the potentially biased small-scale sample. For this reason, robust technical, ethical, and regulatory standards are essential for the responsible implementation of such approaches as well as for unlocking generative AI's potential for BE research and regulatory science.

5. Conclusions

Herein, we propose a new hybrid data augmentation approach for BE studies that combines real clinical data with WGAN-based synthetic data. In the study, we found that hybrid datasets consistently narrow 90% CIs while preserving the primary population characteristics, leading to increased statistical power, and at the design stage of the study, the possibility of obtaining reliable BE assessments with fewer enrolled subjects. The rapid development of generative artificial intelligence highlights the need to take responsible action and seize opportunities, even if the regulatory landscape is still far up the pipe when it comes to approving AI-generated datasets for confirmatory testing. Hybrid datasets are positioned exclusively as design-stage and sensitivity-analysis tools and are not intended to contribute to confirmatory regulatory evidence. In general, although hybrid real–synthetic datasets appear promising approaches to increasing statistical robustness of BE assessments, this information will benefit from explicit validation criteria and setting regulatory standards that can assist in ensuring transparent, clean and reproducible results. The method must be extended to a wider range of drug classes in subsequent research, negative controls for type I error must be integrated, and mechanistic extensions of synthetic data (e.g., PBPK-guided generative models) should be investigated in an effort to ensure ethically responsible and scientifically grounded integration with BE research.

Author Contributions

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

Funding

This research received no external funding.

Data Availability Statement

The datasets that were used for the implementation of WGANs were accessed and obtained from the original publications [22,23]. 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 conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used Gemini to perform language editing and improve readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Appendix A

Table A1. Results of the ANOVA analysis for the assessment of BE concerning the AUC parameter. The table shows the values across all distributions and drugs.

		Lisinopril				Amlodipine				Aceclofenac			
		GMR	**90% CI Lower**	**90% CI Upper**	**CVw**	**GMR**	**90% CI Lower**	**90% CI Upper**	**CVw**	**GMR**	**90% CI Lower**	**90% CI Upper**	**CVw**
Original Sample		0.99	93.64%	105.41%	21.24%	1.00	96.43%	102.87%	11.52%	1.06	98.08%	114.05%	14.49%
		1.07	97.55%	117.02%	22.38%	1.02	97.47%	107.38%	11.80%	1.02	92.35%	113.58%	11.93%
Generated (x Nsample)	1x	1.00	93.98%	106.40%	21.02%	1.00	96.88%	103.22%	11.44%	1.00	95.41%	104.81%	12.01%
	2x	1.03	98.27%	107.74%	21.42%	1.01	98.66%	103.29%	11.94%	1.06	103.33%	109.44%	11.72%
	3x	1.00	95.10%	105.17%	23.47%	1.00	98.20%	101.84%	11.40%	1.00	96.83%	103.27%	11.39%
Hybrid (sample-generated ratio)	50–50	1.03	98.08%	108.93%	21.13%	1.01	98.32%	104.06%	11.48%	1.01	95.91%	106.78%	13.25%
	50–100	1.04	99.91%	108.68%	21.33%	1.01	99.22%	103.62%	11.73%	1.05	101.72%	109.11%	13.10%
	50–150	1.02	97.35%	106.18%	22.36%	1.01	98.78%	102.40%	11.46%	1.01	97.27%	101.04%	12.94%

Table A2. Results of the ANOVA analysis for the assessment of BE concerning the Cmax parameter. The table shows the values across all distributions and drugs.

		Lisinopril				Amlodipine				Aceclofenac			
		GMR	**90% CI Lower**	**90% CI Upper**	**CVw**	**GMR**	**90% CI Lower**	**90% CI Upper**	**CVw**	**GMR**	**90% CI Lower**	**90% CI Upper**	**CVw**
Original Sample		0.99	92.50%	106.31%	25.07%	0.99	96.21%	101.81%	10.05%	1.02	92.74%	111.60%	17.82%
		1.10	99.53%	121.57%	24.66%	0.98	94.69%	101.21%	8.10%	0.94	79.94%	109.78%	18.38%
Generated (x Nsample)	1x	1.00	92.99%	107.54%	25.19%	1.00	98.31%	101.72%	7.88%	1.00	96.22%	103.93%	18.77%
	2x	1.05	99.90%	110.74%	25.22%	0.97	95.64%	98.65%	7.90%	0.99	94.86%	103.36%	18.87%
	3x	1.00	96.02%	104.14%	24.84%	1.00	98.62%	101.40%	7.54%	1.00	96.45%	103.68%	17.10%
Hybrid (sample-generated ratio)	50–50	1.05	98.89%	111.24%	25.13%	0.99	97.20%	100.72%	8.97%	0.97	90.88%	103.07%	18.30%
	50–100	1.07	101.90%	111.87%	25.15%	0.97	95.95%	98.84%	8.98%	0.97	92.74%	101.89%	18.34%
	50–150	1.02	98.58%	106.39%	24.96%	0.99	98.19%	100.77%	8.79%	0.98	94.66%	102.25%	17.46%

Table A3. Results of the ANOVA analysis for the assessment of BE concerning the AS parameter. The table shows the values across all distributions and drugs.

		Lisinopril				Amlodipine				Aceclofenac			
		GMR	**90% CI Lower**	**90% CI Upper**	**CVw**	**GMR**	**90% CI Lower**	**90% CI Upper**	**CVw**	**GMR**	**90% CI Lower**	**90% CI Upper**	**CVw**
Original Sample		0.90	83.08%	98.01%	29.96%	1.00	91.54%	108.20%	30.33%	1.02	93.07%	112.14%	17.95%
		1.02	91.41%	112.88%	26.06%	0.98	84.22%	113.30%	37.24%	0.94	80.88%	110.05%	17.83%
Generated (x Nsample)	1x	1.00	94.22%	106.13%	26.33%	1.00	92.81%	107.75%	36.13%	1.00	96.35%	103.80%	18.32%
	2x	0.97	91.48%	102.13%	24.60%	0.88	80.54%	96.34%	37.01%	1.00	95.51%	103.82%	16.41%
	3x	1.00	95.93%	104.57%	25.51%	1.00	92.80%	107.76%	38.33%	1.00	96.53%	103.59%	17.93%
Hybrid (sample-generated ratio)	50–50	1.01	95.16%	106.74%	28.14%	0.99	91.30%	107.00%	33.23%	0.97	91.38%	103.24%	18.14%
	50–100	0.98	93.60%	103.18%	27.28%	0.91	91.54%	106.34%	33.67%	0.98	93.43%	102.37%	17.18%
	50–150	1.00	96.33%	104.62%	27.73%	0.99	93.09%	106.17%	34.33%	0.99	94.93%	102.32%	17.94%

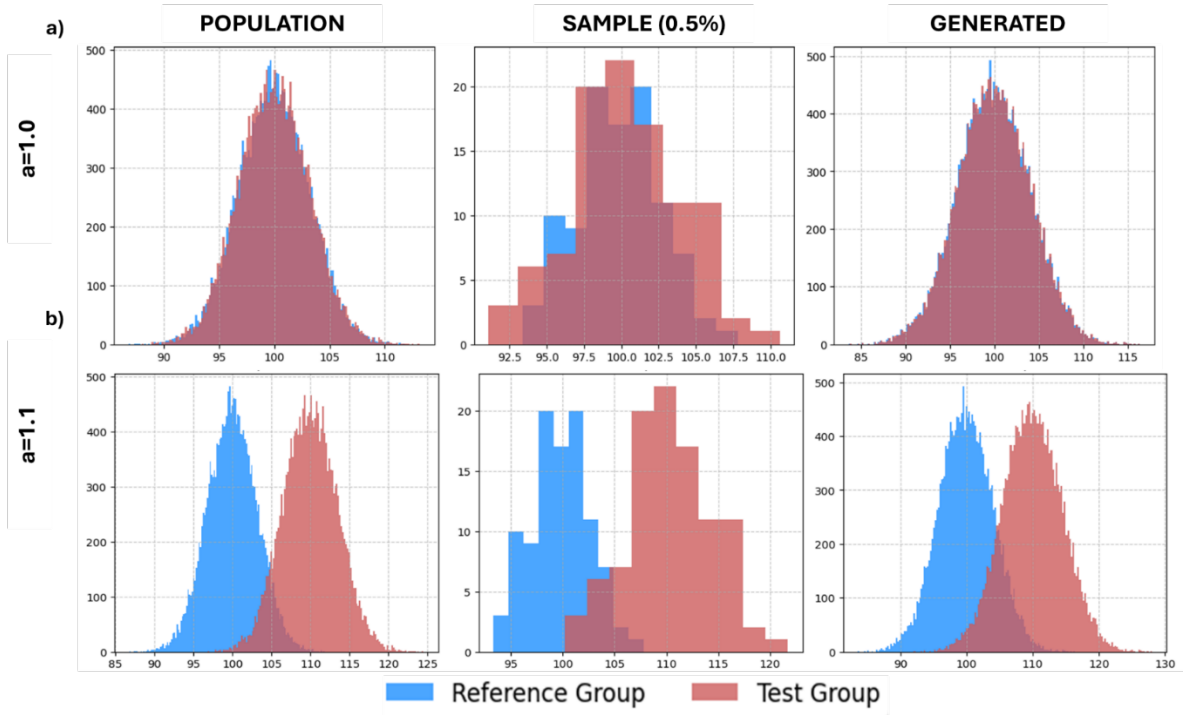

Figure A1. Histograms of distributions from Monte Carlo simulations showcasing the way WGANs work with an initial population of 20,000 subjects. The generated dataset is the same size as the population, the sampling percentage was chosen to be 0.5% (i.e. 100 subjects) and two examples are being presented: (**a**) a = 1.0; and (**b**) 1.1.

References

1. Julious, S.A. *Sample Sizes for Clinical Trials*, 2nd ed.; Chapman and Hall/CRC: New York, NY, USA, 2023. https://doi.org/10.1201/9780429503658.
2. Endrenyi, L.; Tothfalusi, L. Sample Sizes for Designing Bioequivalence Studies for Highly Variable Drugs. *J. Pharm. Pharm. Sci.* **2011**, *15*, 73–84. https://doi.org/10.18433/J3Z88F.
3. Nayyar, A.; Gadhavi, L.; Zaman, N. Machine Learning in Healthcare: Review, Opportunities and Challenges. In *Machine Learning and the Internet of Medical Things in Healthcare*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 23–45. https://doi.org/10.1016/B978-0-12-821229-5.00011-2.
4. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; et al. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144. https://doi.org/10.1145/3422622.
5. Piacentino, E.; Guarner, A.; Angulo, C. Generating Synthetic ECGs Using GANs for Anonymizing Healthcare Data. *Electronics* **2021**, *10*, 389. https://doi.org/10.3390/electronics10040389.
6. Kumar Swarnkar, S.; Guru, A.; Chhabra, G.S.; et al. *Artificial Intelligence Revolutionizing Cancer Care: Precision Diagnosis and Patient-Centric Healthcare*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2024. https://doi.org/10.1201/9781003571339.
7. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.
8. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; et al. Improved Training of Wasserstein GANs. *arXiv* **2017**, arXiv:1704.00028.
9. 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 22 December 2025).
10. U.S. Food and Drug Administration (FDA). Bioavailability and Bioequivalence Studies Submitted in NDAs or INDs—General Considerations. Available online: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioavailability-and-bioequivalence-studies-submitted-ndas-or-inds-general-considerations (accessed on 22 December 2025).
11. Bioequivalence for Immediate-Release Solid Oral Dosage Forms, M13A. Available online: https://database.ich.org/sites/default/files/ICH_M13A_Step4_Final_Guideline_2024_0723.pdf (accessed on 25 March 2026).
12. Farhadi, A.; Taheri, A. Application of GenAI in Synthetic Data Generation in the Healthcare System. In *Application of Generative AI in Healthcare Systems*; Springer: Cham, Switzerland, 2025; pp. 67–89. https://doi.org/10.1007/978-3-031-82963-5_3.
13. Rashidieranjbar, F; Farhadi, A.; Zamanifar, A. Revolutionizing Healthcare with Generative Artificial Intelligence Technologies. In *Generative Artificial Intelligence (AI) Approaches for Industrial Applications*; Springer: Cham, Switzerland, 2025; pp. 189–221. https://doi.org/10.1007/978-3-031-76710-4_10.

14. Farhadi, A.; Zamanifar, A.; Faezipour, M. Application of Generative AI in Drug Discovery. In *Application of Generative AI in Healthcare Systems*; Springer: Cham, Switzerland, 2025; pp. 155–174. https://doi.org/10.1007/978-3-031-82963-5_6.
15. 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.
16. 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. https://doi.org/10.3390/app14114570.
17. Nikolopoulos, A.; Karalis, V.D. Generative Neural Networks for Addressing the Bioequivalence of Highly Variable Drugs. *Algorithms* **2025**, *18*, 266. https://doi.org/10.3390/a18050266.
18. Papadopoulos, D.; Karalis, V.D. Variational Autoencoders for Data Augmentation in Clinical Studies. *Appl. Sci.* **2023**, *13*, 8793. https://doi.org/10.3390/app13158793.
19. Papadopoulos, D.; Karalis, V.D. Introducing an Artificial Neural Network for Virtually Increasing the Sample Size of Bioequivalence Studies. *Appl. Sci.* **2024**, *14*, 2970. https://doi.org/10.3390/app14072970.
20. 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. https://doi.org/10.3390/app14125279.
21. Karalis, V.D. An In Silico Approach toward the Appropriate Absorption Rate Metric in Bioequivalence. *Pharmaceuticals* **2023**, *16*, 725. https://doi.org/10.3390/ph16050725.
22. Li, T.; Liu, Y.; Liu, S.; et al. Bioequivalence Evaluation and Food Effect Assessment of Lisinopril/Amlodipine Tablets in Healthy Chinese Subjects under Fasting and Fed Conditions. *BMC Pharmacol. Toxicol.* **2022**, *23*, 45. https://doi.org/10.1186/s40360-022-00590-6.
23. Bushra, R.; Shoaib, M.H.; Ali, H.; et al. Pharmacokinetics and Bioequivalence Assessment of Optimized Directly Compressible Aceclofenac (100 Mg) Tablet Formulation in Healthy Human Subjects. *PLoS ONE* **2020**, *15*, e0238951. https://doi.org/10.1371/journal.pone.0238951.
24. Karalis, V.D. On the Interplay between Machine Learning, Population Pharmacokinetics, and Bioequivalence to Introduce Average Slope as a New Measure for Absorption Rate. *Appl. Sci.* **2023**, *13*, 2257. https://doi.org/10.3390/app13042257.
25. Shi, Y.; Li, Q.; Zhu, X. Building Footprint Generation Using Improved Generative Adversarial Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 603–607. https://doi.org/10.1109/LGRS.2018.2878486.
26. Bhat, R.; Nanjundegowda, R. A Review on Comparative Analysis of Generative Adversarial Networks' Architectures and Applications. *J. Robot. Control* **2024**, *6*, 53–64. https://doi.org/10.18196/jrc.v6i1.24160.
27. Sengar, S.S.; Hasan, A.B.; Kumar, S.; et al. Generative Artificial Intelligence: A Systematic Review and Applications. *Multimed. Tools Appl.* **2024**, *84*, 23661–23700. https://doi.org/10.1007/s11042-024-20016-1.
28. Fleming, N. Computer-Calculated Compounds: Researchers Are Developing Artificial Intelligence to Discover Drugs. *Nature* **2018**, *557*, S55–S57. https://doi.org/10.1038/d41586-018-05267-X.
29. Beam, A.L.; Kohane, I.S. Big Data and Machine Learning in Health Care. *JAMA* **2018**, *319*, 1317–1318. https://doi.org/10.1001/jama.2017.18391.
30. Committee for Medicinal Products for Human Use (CHMP) Committee for Medicinal Products for Veterinary Use (CVMP). Reflection Paper on the Use of Artificial Intelligence (AI) in the Medicinal Product Lifecycle. Available online: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf (accessed on 25 March 2026).
31. U.S. Food and Drug Administration. Using Artificial Intelligence & Machine Learning in the Development of Drug and Biological Products. Available online: https://www.federalregister.gov/documents/2023/05/11/2023-09985/using-artificial-intelligence-and-machine-learning-in-the-development-of-drug-and-biological (accessed on 25 March 2026).
32. Lucic, M.; Kurach, K.; Michalski, M.; et al. Are GANs Created Equal? A Large-Scale Study. *arXiv* **2017**, arXiv:1711.10337.
33. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. *arXiv* **2018**, arXiv:1812.04948.
34. Brock, A.; Donahue, J.; Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *arXiv* **2018**, arXiv:1809.11096.
35. Salimans, T.; Goodfellow, I.; Zaremba, W.; et al. Improved Techniques for Training GANs. *arXiv* **2016**, arXiv:1606.03498.
36. Frid-Adar, M.; Klang, E.; Amitai, M; et al. Synthetic Data Augmentation Using GAN for Improved Liver Lesion Classification. *arXiv* **2018**, arXiv:1801.02385.
37. Baowaly, M.K.; Lin, C.-C.; Liu, C.-L.; et al. Synthesizing Electronic Health Records Using Improved Generative Adversarial Networks. *J. Am. Med. Inform. Assoc.* **2018**, *26*, 228–241. https://doi.org/10.1093/jamia/ocy142.
38. Choi, E.; Biswal, S; Malin, B.; et al. Generating Multi-Label Discrete Patient Records Using Generative Adversarial Networks. *arXiv* **2017**, arXiv:1703.06490.
39. Papin, J.A.; Mac Gabhann, F.; Sauro, H.M.; et al. Improving Reproducibility in Computational Biology Research. *PLoS Comput. Biol.* **2020**, *16*, e1007881. https://doi.org/10.1371/journal.pcbi.1007881.

40. Hutson, M. Artificial Intelligence Faces Reproducibility Crisis. *Science* **2018**, *359*, 725–726. https://doi.org/10.1126/science.359.6377.725.
41. Mittelstadt, B. Principles Alone Cannot Guarantee Ethical AI. *Nat. Mach. Intell.* **2019**, *1*, 501–507. https://doi.org/10.1038/s42256-019-0114-4.
42. Floridi, L.; Cowls, J. A Unified Framework of Five Principles for AI in Society. *Harv. Data Sci. Rev.* **2019**, 1. https://doi.org/10.1162/99608f92.8cd550d1.
43. Jobin, A.; Ienca, M.; Vayena, E. The Global Landscape of AI Ethics Guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2.
44. Parasrampuria, S.; Sertkaya, A.; Lord, A.; et al. *Cost of Generic Drug Development and Approval*; U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation: Washington, DC, USA, 2021. Available online: https://aspe.hhs.gov/sites/default/files/documents/20e14b66420440b9e726c61d281cc5a5/cost-of-generic-drugs-erg.pdf (accessed on 25 March 2026).
45. Jones, H.; Chen, Y.; Gibson, C.; et al. Physiologically Based Pharmacokinetic Modeling in Drug Discovery and Development: A Pharmaceutical Industry Perspective. *Clin. Pharmacol. Ther.* **2015**, *97*, 247–262. https://doi.org/10.1002/cpt.37.
46. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784.
47. Holzinger, A.; Biemann, C.; Pattichis, C.S.; et al. What Do We Need to Build Explainable AI Systems for the Medical Domain? *arXiv* **2017**, arXiv:1712.09923.