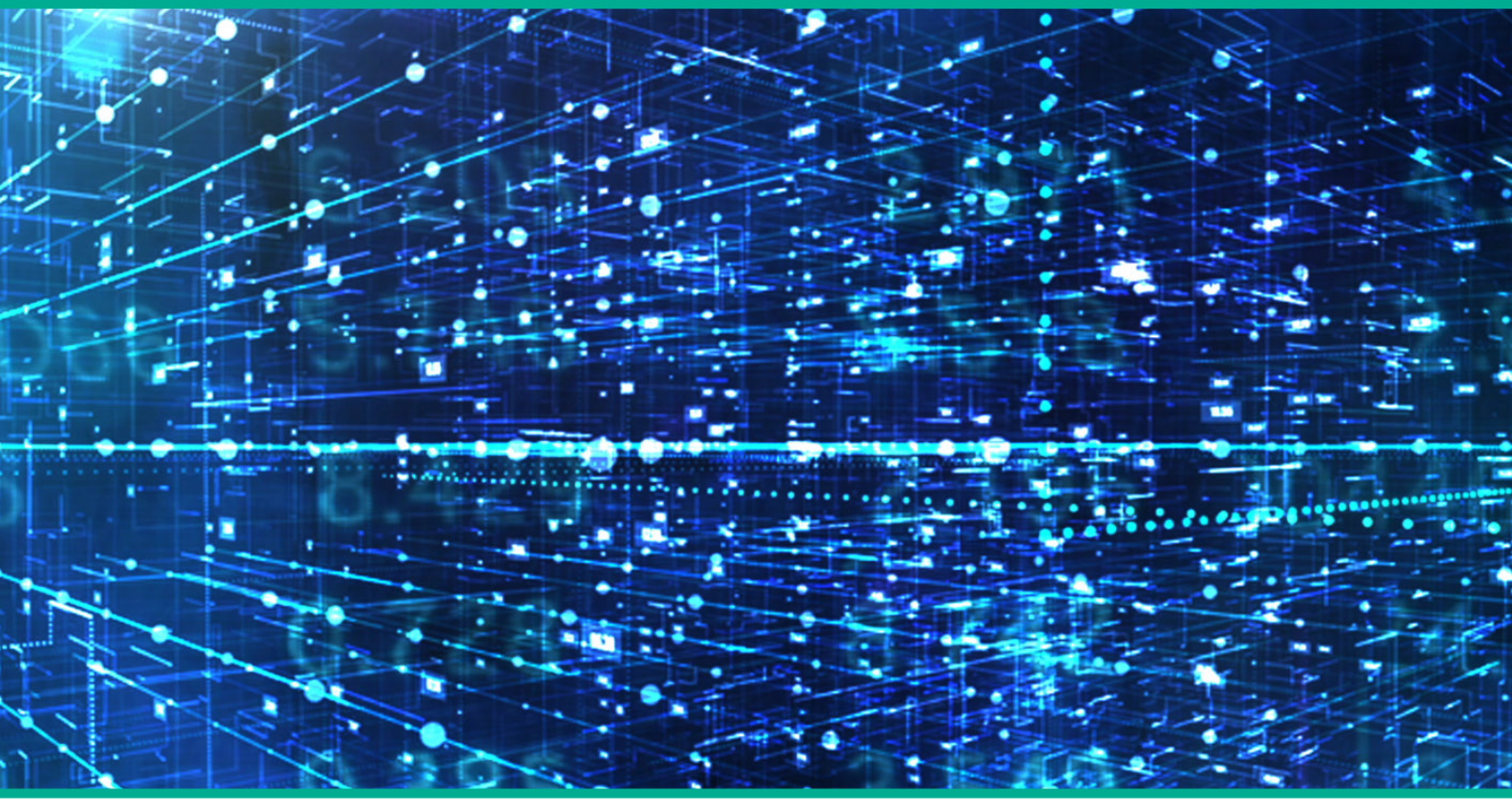
Synthetic data for development of AI as a medical device (AIaMDs)

Regulatory considerations

July 2025



Medicines & Healthcare products Regulatory Agency

PHG FOUNDATION

Authors

Valena Reich, Colin Mitchell, Elizabeth Redrup Hill, Puja Myles, Richard Branson, Russell Pearson and members of the Expert Working Group

About this document

This document presents considerations developed by an expert working group (see appendix) convened by the Medicines and Healthcare products Regulatory Agency and the PHG Foundation.

These considerations do not constitute formal official guidance from the MHRA

This work was supported supported by the Regulators' Pioneer Fund.

URLs in this report were correct as of July 2025

How to reference this report

Reich, V. et al. Synthetic data for development of AI as a medical device (AIaMDs): Regulatory considerations. MHRA & PHG Foundation. 2025.

https://doi.org/10.61599/vhac1

Contents

1. Introduction and scope

This document aims to facilitate dialogue between manufacturers and approving bodies, offering an approach to help the former consider, compile, and justify their use of synthetic data in the development of AI as a Medical Device (AIaMD). It outlines guiding principles for manufacturers, to encourage critical reflection and ensure that the use of synthetic data can be clearly justified as part of the regulatory evidence package.

This document does not constitute official guidance from the MHRA. Rather, it presents considerations developed by an expert working group convened by the MHRA and the PHG Foundation. This work was supported by the Regulators' Pioneer Fund.

For the purposes of this document, *synthetic data refers to artificially generated data that reflects the properties of, and relationships in, real data.* A more detailed definition is included in the glossary. While terminology may vary across disciplines, this document aligns with definitions used in established regulatory frameworks and guidance.

Scope of this document

The scope includes synthetic data generated via data-driven methods, simulation-based approaches, and mechanistic modelling techniques. It also covers the use of synthetic data to boost real-world datasets.

The primary focus is on structured tabular synthetic data used in the pre-market phases of AIaMD development – specifically for training, tuning, and testing AIaMD models for risk prediction or clinical decision support. Synthetic imaging data and synthetic unstructured data (e.g., free text or uncoded data) fall outside this scope, although some of the principles may inform future guidance for such use cases.

While synthetic data may play an important role in AIaMD development, its use in regulatory submissions introduces specific considerations—particularly when it forms a significant part of the evidence package. This document outlines those considerations, building on and complementing existing regulatory guidance. It does not revisit broader topics such as risk classification, clinical evidence, or AIaMD model characteristics, which are already covered in established standards. The sections that follow focus on areas where synthetic data requires particular attention within the regulatory context.

2. Intended use, risk, and justification

Synthetic data should not be considered in isolation, but as part of a broader evidence strategy tailored to the device's intended use and associated risk. This section outlines key contextual factors that influence the regulatory acceptability of synthetic data and suggests an approach to support justification for its use.

2.1. Use case and risk profile

When considering the use of synthetic data for developing AIaMDs, it will be essential to consider the device's use case and risk profile.

The level of evidence and certainty required is generally proportional to risk: higher-risk devices may require stronger justification for use of synthetic data and higher levels of certainty regarding evidence on performance, whereas greater uncertainty may be tolerated for lower-risk devices. Risk-based approaches from existing regulatory frameworks (see resources in Section 4) consider not only uncertainty but also its potential implications for safety and performance. In this context, the use of synthetic data should be carefully aligned with broader risk management strategies.

For higher-risk AIaMDs, there may be a greater need to demonstrate that synthetic data is both high-fidelity and representative of the intended use population, minimising the possibility of introducing bias or uncertainty. A structured risk assessment can help manufacturers reflect on how synthetic data fits into their overall risk management strategy, including any limitations or worst-case scenarios.

Reflective questions:

- What is the intended use case of this device?
- What is the risk profile of the device?
- To what extent does the AIaMD model influence the clinical or operational decision it is intended to support, and what are the potential consequences if that decision is incorrect?
- How does the use of synthetic data for this use case affect your risk management approach?
- Have you anticipated potential worst-case scenarios due to uncertainty linked to the synthetic data?
- Can you justify your approach and demonstrate that you have minimised risks as far as possible, appropriate to the AIaMD use case?

Synthetic data may be especially useful when access to real-world data is limited. However, reliance on it should generally decrease as the associated device risk increases. Not all uncertainty must be resolved pre-market — regulators may accept some evidence gaps if addressed through post-market evidence collection (e.g., in cases where synthetic data is used to enhance representativeness). That said, the actual contribution of synthetic data should be assessed on a case-by-case basis for *improving* evidence quality, not just for reassurance purposes.

These contextual reflections provide a foundation for considering how synthetic data should be positioned within the overall evidence package for regulatory evaluation.

2.2. Positioning synthetic data within the broader evidence package

Synthetic data should be part of a pluralistic evidence strategy. Proportionate and adaptive governance is key, aligning regulatory expectations with the lifecycle stage of a product (e.g., see the *PAGIT* framework – Section 4). Manufacturers should clarify their position, balancing prior knowledge, data quality, and the reasoning behind opportunity costs, feasibility, and patient risk.

When assessing synthetic data as part of the evidence base, explaining why it is appropriate in a given case is essential. While synthetic data can support robust regulatory decision-making, it should be contextualised within existing methodologies, ensuring it supplements rather than replaces other forms of established evidence. In the majority of cases, evidence for regulatory purposes should not be based purely on synthetic data, unless there is a very strong justification to do so. In such exceptional cases, the manufacturer should carefully consider and elucidate the limitations of relying only on synthetic data and implications for use, as part of their regulatory submission. Exposure to unproven therapies further underscores the need for a flexible yet rigorous approach, where the integration of synthetic data must be justified within a proportionate regulatory framework.

Reflective questions:

What is the level of reliance on synthetic data within your broader evidence package?

- **Level 1:** Are you using synthetic data to supplement an already large real-world dataset?
- **Level 2:** Are you incorporating synthetic data alongside expert knowledge or mechanistic modelling to compensate for gaps in real data?
- **Level 3:** Are you relying on interpolation or boosting techniques to generate additional synthetic data where real data is limited?
- **Level 4:** Are you relying primarily on synthetic data, without additional real-world evidence or expert-driven adjustments?

Note: Where synthetic data forms a substantial part of the evidence base, post-market surveillance may be critical to addressing residual uncertainty. See Section 3.5 for further considerations.

At which stage of the product lifecycle is your AIaMD?

- Are you in early-stage development, validation, pre-market approval, or post-market monitoring?
- How does the stage of development impact the sufficiency of synthetic data as part of your evidence package?

How does the explainability of your AIaMD model influence the justification for using synthetic data?

- If the AIaMD model is highly explainable, how does this reduce uncertainty and support a lower requirement for clinical validation?
- If the AIaMD model operates as a black box, how do you justify the sufficiency of synthetic data, and what additional evidence do you provide to ensure robustness across diverse populations and clinical contexts?

2.3. Justifying the use of synthetic data

In line with the use case and risk profile of the AIaMD, a clear and proportionate justification should be provided when including synthetic data in a regulatory submission. The reasons for doing so may vary, including technical needs, ethical concerns, or related to the stage of the product lifecycle. The following reasons can assist in framing this justification.

Technical justifications:

- **Dataset boosting:** Dataset boosting enhances AIaMD model performance by generating synthetic data to address limitations in real-world datasets. It extrapolates patterns to improve generalisation. However, when synthetic data is used for boosting, it must be generated in a way that addresses biases present in the real data used for training the AIaMD to ensure the boosted dataset is representative of the intended use population. Manufacturers should also explain what role dataset boosting plays in their bias mitigation approach.
- **Benchmarking:** Synthetic data can be used to compare AIaMD model performance across different AI architectures under standardised conditions.
- **Testing edge cases and out-of-distribution scenarios:** Edge cases refer to rare, unusual, or extreme scenarios that may not be well-represented in real-world data but are crucial for assessing AIaMD model robustness and reliability. Synthetic data can be generated to simulate such cases, helping estimate AI performance in low frequency but high-risk situations.
- **Quantification of uncertainty:** Synthetic data may support Monte Carlo simulations and bootstrapping to better estimate confidence intervals and variability of AIaMD model predictions.

Ethical and regulatory justifications:

- **Facilitating data access:** Synthetic data may be used when real-world data is unavailable, difficult to collect, or restricted due to ethical or legal restrictions.
- **Addressing gaps in real data:** Helps improve AIaMD model representativeness and generalisability by compensating for missing or underrepresented subgroups in real-world datasets.

Lifecycle justifications:

- **To inform extended use pre-market:** Facilitates pre-market evaluation of the AIaMD model performance for extended or out-of-scope use cases that differ from the original purpose.
- **To inform post-market safety monitoring:** Used to supplement real-world evidence collection and continuously assess AIaMD model reliability over time.

Once the justification for synthetic data use is established, the next step is a structured assessment of its quality and appropriateness.

3. Synthetic data assessment

Taking a structured approach to synthetic data assessment is key to building regulatory confidence. This section supports manufacturers in presenting a clear and well-justified approach to synthetic data use — addressing critical elements such as data provenance, quality, and synthetic data generation model performance to demonstrate that the data is suitable for its intended use. The regulatory acceptability of synthetic data rests on three overarching principles: fidelity, representativeness, and transparency. These are the core qualities that regulatory bodies will be looking for when assessing the suitability of the proposed synthetic data in AIaMD development.

Fidelity refers to how well a synthetic dataset captures the statistical and structural properties of real-world data. High-fidelity synthetic data should accurately reflect clinical relationships, making it useful for AIaMD development. When generating synthetic datasets, there is often a trade-off between fidelity and privacy. In the context of AIaMD regulation, fidelity should be prioritised over the incorporation of privacy measures like censoring, suppression, aggregation or differential privacy during the synthetic data generation process. Privacy requirements should be fulfilled using alternative privacy approaches such as secure access to synthetic data by authorised users, so that fidelity is not compromised.

Representativeness ensures that synthetic data reflects the full diversity of the target intended use population, minimising bias and ensuring generalisability. AIaMD models trained on unrepresentative data risk poor performance in real-world applications, potentially leading to inequitable outcomes. Regulatory frameworks, such as the UK MDR 2002 and the STANDING Together consensus recommendations, emphasise the need for manufacturers to demonstrate that the training data supports generalisable AIaMD models, minimising risks related to bias and population misalignment.

Transparency underpins regulatory trust in synthetic data by requiring clear explanation and justification of the synthetic data generation process, assumptions, and intended use. Manufacturers should provide detailed justifications of why synthetic data was used and how it ensures fidelity and representativeness. Additionally, regulatory guidance, such as MHRA's principles on transparency for machine learning-enabled medical devices, reinforces the need for open disclosure of data provenance and validation methods. In the case of synthetic data, this would include details on how it has been generated.

To put these principles into practice, manufacturers are encouraged to assess their synthetic data using the structured reflective questions set out in the sub-sections that follow. These are designed to support critical evaluation and robust justification at various stages of the synthetic data lifecycle. Together, responses to these questions will help regulatory bodies to evaluate how the principles of fidelity, representativeness, and transparency have been fulfilled.

3.1. Source data provenance and suitability

The credibility of synthetic data depends on the provenance of the source data. Understanding where data originates, how it is curated, and its ethical and regulatory considerations ensures transparency and trust in AIaMD development and validation.

Reflective questions:

- What justifies the selection of your source or ground truth dataset for synthetic data generation?
- Is the source data representative of the intended use population?
- Has the source data been modified prior to use in the synthetic data generation model?
- Have any privacy-enhancing techniques (e.g., suppression, anonymisation, aggregation) been applied to the source data?
  - What are the likely implications for the fidelity and diversity of the resulting synthetic data?

3.2. Synthetic Data Generation (SDG) model assessment

The structure, assumptions, and inputs of the synthetic data generation (SDG) model influence the quality, fidelity, and representativeness of the synthetic data it produces. This section outlines key considerations for evaluating how the SDG process may influence the downstream reliability of the data, which by extension, will influence the reliability of AIaMD models trained or tested with it.

Reflective questions:

- What methods have been used to generate the synthetic data, and which SDG model(s) were applied?
- What are the key assumptions underlying the SDG model?
- What parameters were used in the generation process, and how were they selected or tuned?
- Has the synthetic data been generated specifically for this context (e.g. training or testing of your AIaMD), or has it been repurposed from another use?
- How does the SDG method differ from the AIaMD model it supports, if applicable?
- Does the SDG model follow a purely data-driven approach, a mechanistic (process-driven) approach, or a hybrid approach?
- Has the SDG model been previously validated? If so, in what context?
  - *Note: This is particularly important when synthetic data is used to augment or boost real-world data, especially for underrepresented groups or edge cases.*
- Is the SDG model likely to reinforce or amplify biases present in the source data?
- Have you incorporated any external information—beyond the training dataset—to mitigate bias and ensure sufficient diversity, including representation of edge cases?

3.3. Synthetic data quality and suitability

This section outlines the key considerations for evaluating and ensuring synthetic data quality, including fidelity, bias, consistency with real-world data, and safeguards against overfitting.

Reflective questions:

- Have you assessed whether synthetic data introduces, preserves, or mitigates biases from the source data? If synthetic data is used for bias mitigation, how have you evaluated its suitability for this purpose?
- Have you ensured sufficient fidelity to the data? While fidelity requirements should align with the intended application, in the regulatory contexts considered in this document high-fidelity synthetic data will typically be expected.
- Does your synthetic data generation account for potential shifts in medical practice, patient demographics, or disease trends that may impact its representativeness?
- Have you guarded against overfitting? Does the synthetic data support generalisability, or does it risk overfitting?
- How does the synthetic data compare to real data on a univariate level?
- Have you assessed multivariate relationships to ensure that interactions between variables in the synthetic data align with those in the real data?
- Boosting to augment real-world data: Have you externally validated the boosted synthetic data using available real-world data and human expert review? What oversight measures are in place? Have you planned future data collection to further assess generalisability?
- Boosting when real data is unavailable for validation: If real data is inaccessible, what alternative validation methods have you employed to assess appropriateness and minimise the risk of bias?
- Boosting and generalisation boundaries: How do you define the boundaries between acceptable extrapolation and unsafe synthetic generalisation? What safeguards are in place to avoid generating synthetic data distributions that are not clinically plausible? How is clinical relevance ensured in such cases?
- How is uncertainty within the synthetic dataset assessed (either qualitatively or quantitatively), in terms of data completeness, representativeness, or unknown biases?
- Based on your current evidence package, what level of uncertainty remains in your system, and how does the quality of the synthetic data help reduce it? What is the target level of certainty required for your intended use, and what role does synthetic data play in reaching that level?

3.4. Other considerations when using synthetic data

This section addresses the independence and proper separation of datasets used in the AIaMD lifecycle, especially where synthetic data is involved. Demonstrating that synthetic datasets are used in a way that avoids circularity, overfitting, and recursive degradation is essential to maintaining the credibility and generalisability of the AIaMD.

Reflective questions:

- Is the synthetic dataset sufficiently independent from other datasets used in AIaMD training, based on a qualitative assessment of factors such as data source and generation parameters, in a manner appropriate to the intended use?
- Have you assessed how synthetic datasets differ statistically from real-world clinical data, both at the time of generation and during deployment?
- Have you implemented procedural safeguards to prevent data leakage or inappropriate reuse of synthetic datasets across training, tuning, or validation phases that could lead to overfitting?
- If synthetic data has been used across multiple training cycles, what safeguards are in place to prevent the AIaMD from being affected by model collapse?
    - How have you ensured ongoing generalisability and mitigated the risk of recursive performance degradation?

3.5. Post-market considerations

Under the UK MDR 2002, manufacturers are required to conduct post-market surveillance activities. Where synthetic data has been relied upon during the pre-market phase, additional post-market activities may be necessary to address any remaining uncertainty and sustain regulatory confidence.

Manufacturers are encouraged to:

- Collect post-market data over the 5-year period to re-certification to monitor real-world performance.
- Consider implementing additional post-market measures to address uncertainty remaining due to reliance on synthetic data.

Reflective questions:

- What additional post-market measures have you adopted to address residual uncertainty from synthetic data use?
- How will you collect and analyse post-market data to assess whether the pre-market AIaMD performance based on synthetic data has changed?

4. Wider best practices and supporting resources

In addition to the considerations outlined in this document, manufacturers should ensure that their use of synthetic data aligns with established regulatory and scientific best practices.

Good practice expectations

- Have you followed medical device good practice more generally?
- Have you followed state of the art approaches to synthetic data use and generation?
- Since synthetic data use is not an exception to standard data practices, the same good practice principles that apply to real-world data should be followed. Have you ensured compliance with established best practices for medical data handling?

4.1. Supporting resources

Members of the expert working group recommended the following resources. While not exhaustive, this list highlights widely recognised guidelines and standards that may support compliance, transparency, and methodological rigour when integrating synthetic data into medical device development.

Regulatory frameworks & risk management

- ISO 14971:2019 Medical devices — Application of risk management to medical devices
- ISO 13485:2016+A11:2021 Medical devices — Quality management systems — Requirements for regulatory purposes
- ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management
- Proportionate and Adaptive Governance of Innovative Technologies (PAGIT) Framework

AIaMD regulatory guidance and best practice

- ASME V&V 40 standard 'Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices' (2018)
- ASME V&V 70 standard 'Verification and Validation of Machine Learning' (forthcoming)
- FDA Draft Guidance: 'Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products' (2025)

- IMDRF 'Good machine learning practice for medical device development: Guiding principles'
- MHRA 'Transparency for machine learning-enabled medical devices: guiding principles' (2024)
- STANDING Together consensus recommendations (2025 review)
- TRIPOD+AI statement (2024)

Post-market & surveillance guidance

- MHRA's collection on 'Medical devices: post-market surveillance'
- FDA, Health Canada and MHRA joint guidance on 'Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles' (2023)

5. Glossary

Data Drift: "Refers to the change in the input data distribution a deployed model receives over time, which can cause the model's performance to degrade. This occurs when the properties of the underlying data change. Data drift can affect the accuracy and reliability of predictive models." (FDA Digital Health and Artificial Intelligence Glossary)

Model Collapse: Refers to the progressive degradation of AI model performance when synthetic data generated by earlier model iterations is reused in subsequent training cycles, which has also been referred to as Model Autophagy Disorder (MAD). In the context of AI as a Medical Device (AIaMDs), model collapse poses a risk to reliability, as errors and biases may accumulate across training cycles, leading to overfitting, reduced diversity ('mode collapse'), and performance drift. This can result in biased predictions and a diminished ability to generalise to real-world patient populations. To mitigate model collapse, manufacturers should ensure that synthetic datasets remain representative, diverse, and supplemented with real-world data where possible. Additionally, iterative evaluation is necessary to detect degradation over time and maintain both clinical validity and regulatory trust. See, for instance, papers by Alemohammad et. al. and Shumailov et. al.

Overfitting: "In ML, overfitting occurs when a model learns the training data too thoroughly, capturing not just the fundamental patterns, but also noise or random fluctuations. Such a model might excel on the training data, but struggles to generalize to new, unseen data." (FDA Digital Health and Artificial Intelligence Glossary)

Test Data: "These data are used to characterize the performance of an AI system. These data are never shown to the algorithm during training and are used to estimate the AI model's performance after training. Testing is conducted to generate evidence to establish the performance of an AI system before the system is deployed or marketed. For AI-enabled medical products, test data should be independent of data used for training and tuning." (FDA Digital Health and Artificial Intelligence Glossary)

Training Data: "These data are used by the manufacturer of an AI system in procedures and training algorithms to build an AI model, including to define model weights, connections, and components." (FDA Digital Health and Artificial Intelligence Glossary)

Tuning Data: "These data are typically used by the manufacturer of an AI system to evaluate a small number of trained models. This process involves exploring various aspects, including different architectures or hyperparameters (i.e., parameters used to tune the model for the task). The tuning phase happens before the testing phase of the AI system and is part of the training process. While the AI and ML communities sometimes use the term 'validation' to refer to the tuning data and phase, the FDA will not typically use the word 'validation' in this context due to its specific regulatory definition (see 21 CFR 820.3(z))." (FDA Digital Health and Artificial Intelligence Glossary)

Synthetic Data: "Data that have been created artificially (e.g., through statistical modeling, computer simulation) so that new values and/or data elements are generated. Generally, synthetic data are intended to represent the structure, properties and relationships seen in actual patient data, except that they do not contain any real or specific information about individuals.

For example, in healthcare, synthetic data are artificial data that are intended to mimic the properties and relationships seen in real patient data. Synthetic data are examples that have been partially or fully generated using computational techniques rather than acquired from a human subject by a physical system." (FDA Digital Health and Artificial Intelligence Glossary)

Validation: The process of confirming, through examination and objective evidence, that an AIaMD meets its intended use requirements, ensuring its predictions are accurate, reliable, and generalisable to real-world patient cohorts. Validation assesses model performance, efficacy, and safety, identifying issues such as overfitting and bias that could compromise clinical applicability.

Although the gold standard for external pre-market validation is to test an AIaMD on unseen real-world data, it may be appropriate for manufacturers to use synthetic test datasets - and for regulators to accept this - conditional upon the manufacturer's justification. See also: **Test Data**.

6. Acknowledgements

In November 2022, the UK Government awarded the MHRA a grant of £750,387 via the Regulators' Pioneer Fund to deliver the Synthetic data project, which started in September 2023 and finished delivery in March 2025. The RPF is a grant-based fund to enable UK regulators and local authorities to help create a UK regulatory environment that encourages business innovation and investment. The most recent £12m round was delivered by the Department for Science, Innovation and Technology and concluded on 31st of March 2025.

7. Expert working group

Richard Branson

Operations Lead of Innovative Devices, MHRA

Richard Branson works in the Innovative Devices Division of the MHRA and serves as the project manager of the MHRA's synthetic data research team, combining the Software team, CPRD, and external resource in running projects on synthetic data.

Dominique Chu

Deputy Director of Scientific Data and Insights, MHRA

Dominique Chu recently joined MHRA to lead its Epidemiology and Machine learning teams. Before that, he worked on AI Ethics for HMRC. Prior to his roles in the civil service, he was a Senior Lecturer and Head of the Artificial Intelligence and Data Analytics research group at the School of Computing at the University of Kent.

Professor Alastair Denniston

Professor of Regulatory Science and Innovation, University of Birmingham

Alastair Denniston is Professor of Regulatory Science and Innovation at the University of Birmingham and Honorary Consultant Ophthalmologist at University Hospitals Birmingham NHSFT. He is a leader in the field of Artificial Intelligence (AI) and Digital Health Technologies. With Assoc Prof Xiaoxuan Liu, he leads the AI & Digital Health Group, a research and policy group focused on responsible innovation of AI health technologies. The group's work seeks to ensure AI technologies are safe, effective and equitable, and benefit patients and society.

Professor Alejandro Frangi

Bicentennial Turing Chair in Computational Medicine, University of Manchester

Professor Alejandro Frangi holds the Bicentennial Turing Chair in Computational Medicine at the University of Manchester, with appointments in Computer Science and Health Sciences. He directs the Christabel Pankhurst Institute, driving health technology innovation, and is the Royal Academy of Engineering Chair in Emerging Technologies, focusing on precision computational medicine and in silico trials. His work has been recognised with a Fellowship of the Royal Academy of Engineering (2023) and the IEEE EMBS Technical Achievement Award (2021).

Professor Ibrahim Habli

Professor of Safety-Critical Systems, University of York

Professor Ibrahim Habli, an expert in safety-critical systems at the University of York, specialises in the design and assurance of software-intensive systems, particularly in AI and autonomous applications. He leads the UKRI Centre for Doctoral Training in Safe AI Systems (SAINTS) and is Research Director of the Centre for Assuring Autonomy (CfAA). Collaborating with ethicists, clinicians, and industry leaders like Jaguar Land Rover and NASA, he contributes to advancing AI safety. He also advises on safety standards through committees such as BSI DS/1, EUROCAE, and IEEE.

Dr Marinos Ioannides

Head of Software, MHRA

Dr Marinos Ioannides is an experienced professional with a background in medicine, data science, and public service. Having worked in clinical medicine, he witnessed firsthand how poor data and methods can lead to harmful decisions. As a former researcher at Cochrane, he saw how flawed methodologies are sometimes intentionally used to skew findings. In his civil service career, including as Head of Data Science at the Cabinet Office, he observed innovative methods sidelined for being 'too complex'. Now at the Medicines and Healthcare products Regulatory Agency (MHRA), Marinos is committed to addressing these challenges.

Dr Puja Myles

Director, Clinical Practice Research Datalink, MHRA

Dr Puja Myles is Director of the UK Medicines and Healthcare products Regulatory Agency's (MHRA) specialist real world data research service, Clinical Practice Research Datalink (CPRD). She trained as a public health specialist and epidemiologist following her initial clinical training and practice. She was a public health academic working with real world data for many years before she joined CPRD. She is a fellow of the Faculty of Public Health (UK), a senior fellow of the Higher Education Academy (UK) and holds a PhD in Epidemiology. Her current research areas include real world data research, data quality, synthetic data and AI explainability.

Johan Ordish

Global Head of Digital Health and Innovation Policy, Roche

Johan Ordish is Global Head of Medical Device Policy for Apple Health. Prior to Apple, Johan was the Global Head of Digital Health and Innovation Policy at Roche. He also co-chairs the International Medical Device Regulators Forum's Generative AI Sub Group and was a co-author of Good Machine Learning Practice for Medical Device Development. Previous to Roche, Johan was Head of Software and AI at the Medicines and Healthcare products Regulatory Agency (MHRA), leading work on the regulation of software as a medical device for the UK. Johan is also an Honorary Associate Professor at the College of Medical and Dental Sciences, University of Birmingham and a By-Fellow with Hughes Hall, University of Cambridge.

Dr Mark Palmer

Senior Chief Technologist for Healthcare, Ansys

Dr Mark Palmer serves as the Senior Chief Technologist for Healthcare at Ansys, where he leads initiatives to integrate computational modelling and simulation into the healthcare sector. He focuses on creating personalized patient models to evaluate tailored interventions and advancing simulation tools to support collaboration among diverse healthcare experts. Dr Palmer also drives innovations in digital evidence for regulatory approval, streamlining model preparation, data automation, and knowledge sharing to enable safer and more efficient clinical studies.

Dr Russell Pearson

AI Regulatory Policy and Projects Lead, MHRA

For the past 12 years Dr Russell Pearson has worked both in the public and private sector with a focus on bringing innovative medical device technologies to the regulated healthcare market. In his current role at the MHRA, Russell focuses on the regulation of Artificial Intelligence as a Medical Device (AIaMD) for the UK market, the MHRAs regulatory sandbox for AI and alignment with other medical device regulators internationally through chairing the International Medical Device Regulators Forum (IMDRF) AI working group.

Dr Gavin Quigley

Internal Clinician in Global Regulatory Compliance at BSI

Dr Gavin Quigley, an Internal Clinician in Global Regulatory Compliance at BSI, leverages 16 years of neurosurgery experience and an MBA to ensure the safety and quality of medical devices. With expertise in intraoperative imaging, clinical trials, and medical education, he combines clinical knowledge and business acumen to deliver innovative solutions. Passionate about collaboration, he works with multidisciplinary teams to address complex healthcare challenges and broaden his impact beyond traditional clinical practice.

Professor Allan Tucker

Head of Intelligent Data Analysis Group, Brunel University

Allan Tucker, Professor of Artificial Intelligence and head of the Intelligent Data Analysis Group at Brunel University, has over 21 years of research experience and 120 peer-reviewed publications. His work includes collaborations with Google, Moorfields Eye Hospital, the Royal Free Hospital, and the MHRA, focusing on AI benchmarking, explainability, and in-silico trials. He has led multiple Innovate UK Pioneer Funds and contributes to leading AI conferences and journals. His research is renowned for its ethical and practical impact on health and medical fields.

Professor Christopher Yau

Professor of Artificial Intelligence, Big Data Institute, Oxford University

Professor of Artificial Intelligence at the Big Data Institute, University of Oxford, he works across the Nuffield Departments of Women's and Reproductive Health and Population Health. His research leverages high-dimensional molecular and healthcare data to study human diseases, particularly cancer, and predict health outcomes. He collaborates with the CONSORT/SPIRIT-AI consortium and the MHRA to develop guidelines for AI-based medical devices. A Turing AI Fellow, he also leads the HDR UK-Turing Wellcome PhD Programme in Health Data Science.

PHG Foundation

Dr Colin Mitchell

Head of Humanities, PHG Foundation

Dr Colin Mitchell is the Head of Humanities at the PHG Foundation, leading work on legal and ethical issues raised by new health technologies, genomics, biomedical research, and novel data approaches. Colin has a background is as an academic legal researcher at the universities of Amsterdam and Oxford and expertise in the interface between health policy, scientific research and the law.

Dr Elizabeth Redrup Hill

Senior Policy Analyst in Law and Regulation, PHG Foundation

Dr Elizabeth Redrup Hill is a Senior Policy Analyst within the Humanities Team at the PHG Foundation. She works on and has expertise in the legal and ethical challenges raised by novel health innovations and data approaches. Elizabeth has taught medical law and ethics and criminal law at the universities of Southampton and Imperial College London.

Valena Reich

Policy Analyst in Humanities, PHG Foundation

Valena Reich is a Policy Analyst at the PHG Foundation, focusing on ethical, legal, and societal challenges in health innovation. She has an academic background in philosophy and the ethics of AI.

Medicines & Healthcare products Regulatory Agency

The Medicines and Healthcare products Regulatory Agency (MHRA) regulates medicines, medical devices and blood components for transfusion in the UK. The MHRA is responsible for making sure these products meet set standards for safety, quality and effectiveness.

The MHRA is an Executive Agency of the Department of Health and Social Care.

www.gov.uk/mhra

PHG FOUNDATION

The PHG Foundation is a health policy think-tank and linked exempt charity of the University of Cambridge. We work to achieve better health through the responsible and evidence-based application of biomedical science. We are a registered company, no. 5823194.