DPLONGSYN: DIFFERENTIALLY PRIVATE LONGITUDINAL IMAGE SYNTHESIS

Anonymous authors
Paper under double-blind review

ABSTRACT

Longitudinal image data capture changes in the same subject over time, such as anatomical changes across medical scans of the same patient. Such data often contain sensitive individual information, motivating synthetic data release with formal differential privacy (DP) guarantees. However, existing DP image synthesis methods mainly focus on individual images, while longitudinal synthesis also requires coherent changes between observations that agree with the associated records. To this end, we introduce DPLongSyn, a DP longitudinal image synthesis framework that learns record-conditioned longitudinal transitions and uses them to synthesize follow-up images from public source images. DPLongSyn first learns a transition prior from public longitudinal pairs. It then uses private data to generate DP synthetic records describing subject attributes and longitudinal changes and to adapt the transition model using DP-SGD. At synthesis time, the adapted model generates a deformation conditioned on a public source image and a DP synthetic record, then applies it to produce the follow-up image. Experiments on three benchmarks show that DPLongSyn achieves better synthetic data quality, longitudinal consistency, and downstream utility than DP synthesis baselines. To the best of our knowledge, this is the first work to study DP longitudinal image synthesis.

1 INTRODUCTION

Longitudinal image data capture how the same subject changes over time, such as anatomical progression in medical scans (Fischer et al., 2025), changes in facial appearance (Wen et al., 2015), or temporal changes in observed scenes (Tewkesbury et al., 2015). By linking observations, these data reveal changes that an isolated image cannot establish and support learning about progression and temporal dynamics. However, these data often contain sensitive individual information, raising privacy concerns when shared (Fischer et al., 2025; Giouroukou et al., 2025). This motivates synthetic releases that retain longitudinal information while protecting the private data used to learn it.

Differentially private (DP) synthetic data has emerged as a widely studied approach for releasing useful data with formal privacy guarantees (Dwork et al., 2006; Zhang et al., 2021; Lin et al., 2024). However, existing DP image synthesis methods mainly focus on generating individual images (Chen et al., 2026b; Gong et al., 2025). For longitudinal synthesis, realistic individual images are insufficient: two plausible brain scans may form an implausible progression. For image pairs accompanied by records describing subject attributes and longitudinal changes, the generated changes must also agree with those records. Even a plausible image transition can provide a misleading training example if it contradicts its record. This leads to the central question of this work:

Can we synthesize longitudinal images under DP while preserving meaningful transitions?

To address this question, we introduce **DPLongSyn**, a new method for synthesizing longitudinal image pairs and associated records with a DP guarantee (Figure 1). Because longitudinal observations share substantial visual content, we use public data to supply source appearance and a pretrained transition representation, while privately learning the record distribution and adapting the transition model. Record conditioning connects these components: a DP synthetic record guides the generation of a deformation that transforms a public source image into a follow-up.

Our main contributions are summarized as follows:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

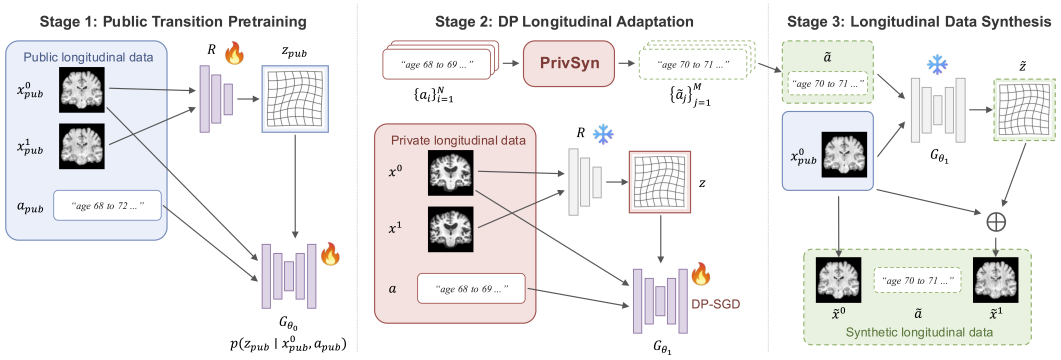


Figure 1: Overview of DPLongSyn. DPLongSyn first learns a registration model R and a conditional diffusion model G_{θ_0} from public longitudinal data. It then generates DP synthetic records with PrivSyn and adapts the diffusion model to private longitudinal changes using DP-SGD, while keeping R fixed. At synthesis time, the adapted model generates a transition $\tilde{z}$ conditioned on a DP synthetic record $\tilde{a}$ and a public source image $\tilde{x}^0$, which is used to produce the synthetic follow-up $\tilde{x}^1$. The final release consists of $(\tilde{x}^0, \tilde{a}, \tilde{x}^1)$.

- We introduce DPLongSyn, which synthesizes longitudinal image pairs and associated records under DP by explicitly learning record-conditioned transitions. It combines DP record synthesis with parameter-efficient private adaptation of a publicly pretrained transition model, using public images as sources.
- We establish an (ϵ, δ)-DP guarantee for DPLongSyn by composing the privacy costs of record synthesis and transition adaptation. Once these components are fixed, synthetic data can be generated using only public data and DP outputs, without further privacy loss.
- We evaluate DPLongSyn on two brain MRI datasets and one facial-expression dataset across multiple privacy levels. DPLongSyn improves synthetic data quality and longitudinal consistency, and achieves stronger overall performance on follow-up and longitudinal progression classification tasks than existing DP synthesis baselines.

2 Related Work

Differentially Private Image and Multimodal Synthesis. Prior work on DP image synthesis has mainly focused on generating individual images, including feature-based methods (Harder et al., 2021), GAN-based approaches (Chen et al., 2020), and diffusion-based generative models (Dockhorn et al., 2023; Gong et al., 2025). Moving beyond image-only synthesis, Chen et al. (2026b) study DP synthesis of images paired with multivariate records, and explore several ways to model the dependence between these modalities. DP-Indep synthesizes the two modalities independently using AIM and DPDM (McKenna et al., 2022; Dockhorn et al., 2023); DP-JointMERF and DP-JointDM model them jointly using private feature matching and diffusion models, respectively (Harder et al., 2021; Dockhorn et al., 2023; Kotelnikov et al., 2023); and DP-ImageTab and DP-TabImage use conditional factorizations to generate one modality from the other (Chen et al., 2026b). We use these methods as baselines because they synthesize both images and associated records under DP. In our longitudinal adaptations, each source–follow-up pair is represented as a multichannel image, allowing the image generators to learn dependencies between observations. For DP-Indep, independence is between the image pair and its record, not between the observations. DPLongSyn explicitly generates a record-conditioned deformation and applies it to a public source image.

Longitudinal Image Synthesis. Longitudinal image synthesis has been studied across a range of applications, including modeling changes in appearance, anatomy, aging, and disease progression over time (Nhan Duong et al., 2016; Duong et al., 2019; Li et al., 2021). Recent methods incorporate diffusion models, temporal modeling, and structured attributes to improve individualized and temporally consistent follow-up synthesis (Puglisi et al., 2025; Chen et al., 2026a). A related line of work explicitly models geometric changes through deformation fields or continuous shape dynamics rather than directly synthesizing image intensities (Kim et al., 2022; Park et al., 2026; Wu et al., 2025). These methods do not provide DP guarantees for their training data. DPLongSyn addresses

this gap by providing DP guarantees for longitudinal image synthesis, using IGG's deformation-based generative framework (Wu et al., 2025) as a publicly pretrained component.

3 Problem Formulation

We study DP synthesis of longitudinal samples containing two ordered image observations and an associated structured record.

Definition 3.1 (Longitudinal Image-Record Dataset)**.** We define a longitudinal image-record dataset $\mathcal{D}$ as a collection of N tuples, denoted by $\mathcal{D} = \left\{(x_i^0, a_i, x_i^1)\right\}_{i=1}^{N}$, where x_i^0 and x_i^1 are two ordered image observations of the same individual, and a_i is the associated fixed-dimensional record containing attributes of the individual and the change between the two observations.

Examples include two scans of the same patient with anatomical-change records, or two facial observations with expression-change records. We allow auxiliary public data $\mathcal{D}_{\text{pub}}$, which are held fixed when comparing neighboring private datasets.

Definition 3.2 (DP Longitudinal Image Synthesis)**.** Given a sensitive longitudinal dataset $\mathcal{D}$, the goal of DP longitudinal image synthesis is to design a randomized mechanism $\mathcal{A}(\mathcal{D}; \mathcal{D}_{\text{pub}})$ that generates a synthetic dataset $\tilde{\mathcal{D}} = \left\{(\tilde{x}_j^0, \tilde{a}_j, \tilde{x}_j^1)\right\}_{j=1}^{M}$ while satisfying (ϵ, δ)-DP. The output size M is prescribed independently of the private data. For any neighboring datasets $\mathcal{D}$ and $\mathcal{D}'$ that differ in one complete image–record–image tuple, and any measurable set $\mathcal{S}$ of possible outputs,

$$\mathbb{P}[\mathcal{A}(\mathcal{D}; \mathcal{D}_{\text{pub}}) \in \mathcal{S}] \leq e^{\epsilon}\mathbb{P}[\mathcal{A}(\mathcal{D}'; \mathcal{D}_{\text{pub}}) \in \mathcal{S}] + \delta.$$

The synthetic dataset notation describes the released tuples, without requiring every image to be generated from scratch. In DPLongSyn, $\tilde{x}_j^0$ is selected from public images, while $\tilde{a}_j$ and $\tilde{x}_j^1$ are synthesized. A useful release additionally requires image fidelity, record fidelity, and agreement between recorded and image-based changes. These utility properties are evaluated separately from the privacy guarantee.

4 DPLongSyn

4.1 Motivation and Design Principles

Before presenting DPLongSyn, we first motivate its design from the structure of longitudinal data.

Longitudinal observations of the same subject share most of their visual content. We examine this property using longitudinal brain MRI pairs from ADNI, comparing image differences within and across subjects. As shown in Figure 2, within-subject pairs are substantially more similar: the median $1 - \text{SSIM}$ is 0.21, compared with 0.51 for pairs from different subjects. This suggests that most visual content persists across observations, and the transition between them is relatively small. A generator that synthesizes both observations must produce their shared appearance as well as their differences. We instead take the source image from public data and privately learn record-conditioned transitions for generating the follow-up. This design assumes access to a public reference dataset in a compatible image domain. It provides source appearance and a general transition prior without requiring the public cohort to match the private target population.

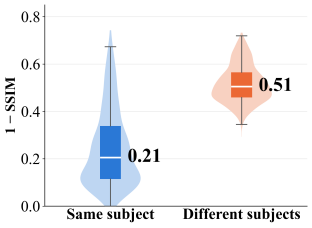


Figure 2: Longitudinal pairs are more similar.

With x^0 supplied by public data, the private learning problem has two remaining parts. A longitudinal sample contains an associated record a and a transition from a source image x^0 to a follow-up image x^1. A DP synthesis method must preserve both the distribution of records and how the image transition depends on the associated record. DPLongSyn therefore uses two DP components: record synthesis to model $p(a)$ and transition learning to model $p(x^1 \mid x^0, a)$.

This motivates the design of **DPLongSyn**: public data provides source images, while private data is used to learn the record distribution and how transitions depend on records. Conceptually, DP-

LongSyn factorizes the generated distribution as

$$p_{\text{DPLongSyn}}(x^0, a, x^1) = p_{\text{DP}}(a)\, q_{\text{pub}}(x^0 \mid a)\, p_{\text{DP}}(x^1 \mid x^0, a).$$

Here, $p_{\text{DP}}(a)$ models the DP synthetic record distribution, q_{pub} selects a public source image, and $p_{\text{DP}}(x^1 \mid x^0, a)$ denotes the DP learned transition from the source image to the follow-up image.

4.2 DPLongSyn Algorithm

We now introduce the DPLongSyn algorithm, which consists of three stages. First, we use a public longitudinal dataset $\mathcal{D}_{\text{pub}}$ to learn a general transition prior. Second, given a sensitive longitudinal dataset $\mathcal{D}$, we synthesize longitudinal records under DP and privately adapt the transition model. Finally, we combine a synthetic record with a public source image and sample a corresponding transition to generate the follow-up image.

Stage 1: Public Transition Pretraining. We build on the deformation-based generative framework of IGG (Wu et al., 2025) to model longitudinal transitions. We first train a registration model R on the public longitudinal dataset $\mathcal{D}_{\text{pub}}$. For each public longitudinal pair $(x^0_{\text{pub}}, a_{\text{pub}}, x^1_{\text{pub}})$, R encodes the transformation from the source image x^0_{pub} to the follow-up image x^1_{pub} as a latent deformation representation,

$$z = R(x^0_{\text{pub}}, x^1_{\text{pub}}).$$

The latent representation z can be decoded into a deformation field that maps the source image to the follow-up image. We defer details of the deformation representation and registration procedure to Appendix A.1.

We then train a conditional latent diffusion model G_{θ_0} (Rombach et al., 2022) on $\mathcal{D}_{\text{pub}}$ to model

$$p(z \mid x^0_{\text{pub}}, a_{\text{pub}}).$$

The resulting model serves as a general transition prior that is later adapted to the private longitudinal data. Since both R and G_{θ_0} are trained on $\mathcal{D}_{\text{pub}}$, this stage incurs no privacy cost.

Stage 2: DP Longitudinal Adaptation. We use the private dataset to model two complementary components under DP: the longitudinal record distribution and the record-conditioned transition distribution. The former is modeled through DP record synthesis, while the latter is learned by privately fine-tuning the conditional latent diffusion model.

DP Record Synthesis. Let $\mathcal{A} = \{a_i\}_{i=1}^N$ denote the record component of the sensitive dataset $\mathcal{D}$. We use PrivSyn (Zhang et al., 2021), a tabular synthesis method that satisfies ρ_0-zero-concentrated differential privacy (zCDP) (Bun & Steinke, 2016), to generate a synthetic record dataset

$$\tilde{\mathcal{A}} = \textsc{PrivSyn}(\mathcal{A}, \rho_0).$$

PrivSyn synthesizes tabular data by privately estimating low-order marginals and generating records that approximately preserve these distributions. Further details are provided in Appendix A.2.

Parameter-Efficient DP Transition Adaptation. For each private longitudinal pair (x^0_i, a_i, x^1_i), the fixed public registration model R extracts its latent transition representation,

$$z_i = R(x^0_i, x^1_i).$$

Starting from the publicly pretrained model G_{θ_0}, we privately fine-tune the conditional latent diffusion model on $\{(z_i, x^0_i, a_i)\}_{i=1}^N$ using DP-SGD (Abadi et al., 2016), so that it adapts to the private latent transitions conditioned on the source image and associated record.

Public pretraining provides a transition representation, but both the transition distribution and its relationship with the associated records may differ in the private population. We therefore restrict private adaptation to a small subset of parameters while retaining the remaining publicly pretrained weights. This design aims to balance adaptation capacity with preservation of the public representation under DP training. The dataset-specific parameter choices are described in Section 5. The resulting model G_{θ_1} satisfies $(\alpha, \gamma_1(\alpha))$-Rényi differential privacy (RDP) (Mironov, 2017).

Together, PrivSyn and G_{θ_1} model the longitudinal record distribution $p(a)$ and the record-conditioned transition distribution $p(z \mid x^0, a)$, respectively, under DP.

Stage 3: Longitudinal Data Synthesis. For each synthetic record $\tilde{a}_j$ sampled from $\tilde{\mathcal{A}}$, DPLongSyn selects a source image $\tilde{x}_j^0$ from the public dataset $\mathcal{D}_{\text{pub}}$. If the public images are not well matched to the baseline characteristics of the synthetic records, we use record-guided matching to select a source whose baseline attributes are consistent with $\tilde{a}_j$. Otherwise, the source is sampled directly from $\mathcal{D}_{\text{pub}}$. Conditioned on the selected source image $\tilde{x}_j^0$ and synthetic record $\tilde{a}_j$, we sample a latent transition from the DP-fine-tuned conditional latent diffusion model G_{θ_1},

$$\tilde{z}_j \sim G_{\theta_1}(\cdot \mid \tilde{x}_j^0, \tilde{a}_j).$$

$\tilde{z}_j$ is then decoded into the deformation field and applied to $\tilde{x}_j^0$ to generate the synthetic follow-up image $\tilde{x}_j^1$. Repeating this process produces the synthetic dataset $\tilde{\mathcal{D}} = \left\{(\tilde{x}_j^0, \tilde{a}_j, \tilde{x}_j^1)\right\}_{j=1}^{M}$.

All operations in this stage depend only on public data or outputs of the DP mechanisms and therefore incur no additional privacy loss by the post-processing property of DP.

The complete DPLongSyn procedure is summarized in Algorithm 1.

Algorithm 1 DPLongSyn

Input: private dataset $\mathcal{D} = \{(x_i^0, a_i, x_i^1)\}_{i=1}^N$, public dataset $\mathcal{D}_{\text{pub}}$, privacy target (ϵ, δ), number of synthetic samples M
Train the registration model R and conditional latent diffusion model G_{θ_0} on $\mathcal{D}_{\text{pub}}$
$\{\tilde{a}_j\}_{j=1}^M \leftarrow \textsc{PrivSyn}(\{a_i\}_{i=1}^N, \rho_0)$
$z_i \leftarrow R(x_i^0, x_i^1)$ for $i = 1, \ldots, N$
Fine-tune the selected parameters of G_{θ_0} with DP-SGD to obtain G_{θ_1} with $(\alpha, \gamma_1(\alpha))$-RDP.
$\tilde{\mathcal{D}} \leftarrow \emptyset$
for $j = 1, \ldots, M$ **do**
    Select a public source $\tilde{x}_j^0 \in \mathcal{D}_{\text{pub}}$ for $\tilde{a}_j$
    $\tilde{z}_j \sim G_{\theta_1}(\cdot \mid \tilde{x}_j^0, \tilde{a}_j)$
    Decode $\tilde{z}_j$ into a deformation and apply it to $\tilde{x}_j^0$ to obtain $\tilde{x}_j^1$
    $\tilde{\mathcal{D}} \leftarrow \tilde{\mathcal{D}} \cup \{(\tilde{x}_j^0, \tilde{a}_j, \tilde{x}_j^1)\}$
end for
return $\tilde{\mathcal{D}}$

4.3 Privacy Guarantee

Theorem 4.1. *DPLongSyn satisfies (ϵ, δ)-DP with respect to the sensitive longitudinal dataset $\mathcal{D}$.*

Proof. The privacy loss of DPLongSyn comes from two mechanisms: PrivSyn for DP record synthesis and DP-SGD for private adaptation of the conditional latent diffusion model. PrivSyn satisfies ρ_0-zCDP, while the DP-SGD adaptation satisfies $(\alpha, \gamma_1(\alpha))$-RDP. We compose the two mechanisms in RDP and convert the resulting guarantee to (ϵ, δ)-DP. All subsequent synthesis steps depend only on public data or outputs of these DP mechanisms and therefore incur no additional privacy loss by the post-processing property of DP. The privacy accounting details are provided in Appendix A.5. □

5 Experiments

Datasets. We evaluate DPLongSyn on three settings across two domains: brain MRI and facial expressions. For brain MRI, we use the Dallas Lifespan Brain Study (DLBS) (Park et al., 2025) as public data, and ADNI (Weiner et al., 2010) and OASIS-3 (LaMontagne et al., 2019) as two separate private datasets. Each sample contains two scans of the same subject at different visits and a record describing the diagnosis transition, age at the first visit, time between visits, and changes in hippocampal and ventricular volumes. DLBS contains cognitively normal adults, while ADNI and OASIS-3 span normal aging and cognitive decline. For facial expressions, we use KDEF (Lundqvist et al., 1998) as public data and CelebV-HQ (Zhu et al., 2022) as private data. Each sample contains a pair of face images and a record describing changes in facial geometry, including mouth opening, mouth width, brow raise, and eye opening. KDEF contains facial expressions captured under

controlled conditions, while CelebV-HQ contains Internet face videos that we process into paired images and expression-change records. Dataset preprocessing details are provided in Appendix B.2.

Setup for DPLongSyn. For brain MRI, we encode the structured numerical record, including age, diagnosis, and anatomical changes, using an MLP. For facial expressions, we encode the categorical expression record using a pretrained BERT encoder (Devlin et al., 2019). During private adaptation, we update normalization parameters and the record-conditioning MLP for brain MRI; for facial expressions, we update only normalization parameters and keep BERT frozen. We evaluate $\epsilon \in \{2, 8, 32\}$ with $\delta = 10^{-5}$. For each setting and privacy level, we generate $M = 2{,}000$ synthetic samples, as for all baselines. Additional implementation details are provided in Appendix B.4.

Evaluation. We evaluate synthetic data quality in terms of image fidelity, record fidelity, and longitudinal consistency. For image fidelity, we report Fréchet Inception Distance (FID) (Heusel et al., 2017) and Kernel Inception Distance (KID) (Bińkowski et al., 2018) between synthetic and held-out real follow-up images. For record fidelity, we report the mean per-attribute total variation distance (TVD) between synthetic and real training records. Lower FID, KID, and TVD indicate better fidelity. For longitudinal consistency, we use Spearman's rank correlation ρ (Spearman, 1961) between the change measured from each generated image pair and the change specified in its record. For brain MRI, we compare measured ventricular change with the recorded ventricular-volume change (Appendix B.7). For facial expressions, we compute this correlation for mouth opening, mouth width, brow raise, and eye opening, and average across the four attributes. Higher correlation indicates better consistency. Metric implementation details are provided in Appendix B.3.

We further evaluate synthetic data utility on two downstream tasks: follow-up classification and longitudinal progression classification.

(1) For follow-up classification, we train the same set of 24 classifiers on each synthetic release and evaluate them on the held-out private test set. Eighteen classifiers combine frozen features from DINOv2 ViT-B/14 (Oquab et al., 2023), ImageNet-pretrained ResNet-50 (He et al., 2016), or ImageNet-pretrained Swin-T (Liu et al., 2021) with one of six prediction heads: logistic regression, linear SVM, random forest, XGBoost, k-nearest neighbors, or an MLP. The remaining six are end-to-end models: ResNet-18 (He et al., 2016), DenseNet-121 (Huang et al., 2017), EfficientNet-B0 (Tan & Le, 2019), ViT-S/16 (Dosovitskiy, 2020), DeiT-S/16 (Touvron et al., 2021), and Swin-T (Liu et al., 2021). We report macro AUC and balanced accuracy averaged across the 24 classifiers.

(2) For longitudinal progression classification, we evaluate whether synthetic image pairs preserve information about disease progression. On ADNI, we consider subjects diagnosed with MCI at the baseline visit and predict whether they progress to Dementia at the follow-up visit. We use the same 24 classifier architectures, adapted to take both visits as input. For the frozen-feature models, we concatenate the baseline feature, follow-up feature, and their difference, $[F(x^0), F(x^1), F(x^1) - F(x^0)]$. For the end-to-end models, a shared backbone encodes both visits, and the two features and their difference are passed to the classification head. We report AUC averaged across the 24 classifiers. Additional evaluation details are provided in Appendix B.6.

Baselines. We compare DPLongSyn with five DP paired table–image synthesis methods from Chen et al. (2026b): DP-TabImage, DP-ImageTab, DP-JointDM, DP-Indep, and DP-JointMERF. They are the closest existing baselines to our setting because they jointly synthesize images and the associated records under DP, although they do not model longitudinal transitions. For a fair comparison, all methods use the same public data for pretraining, the same privacy levels, and downstream evaluation protocol as DPLongSyn. Adaptations required for our datasets are described in Appendix B.5.

5.1 DP Synthetic Data Quality

In this section, we evaluate DPLongSyn against DP paired image–record synthesis baselines on three datasets at privacy budgets of $\epsilon \in \{2, 8, 32\}$. The results are presented in Table 1. Overall, DPLongSyn achieves the best result, with improvements spanning image fidelity, record fidelity, and longitudinal consistency. For example, at $\epsilon = 8$, DPLongSyn achieves the best result on every metric across all three datasets. On ADNI, it reduces FID from 137.7 for the strongest baseline to 93.8 and increases longitudinal correlation from 0.036 to 0.161. On CelebV-HQ, the corresponding correlation increases from 0.041 to 0.169.

Table 1: DP synthetic data quality. Image fidelity is measured by FID and KID, record fidelity by TVD, and longitudinal consistency by Spearman's rank correlation ρ between the recorded and measured changes. "–" indicates that the correlation is undefined because no valid measurable structure is available.

	Method	ADNI				OASIS-3				CelebV-HQ			
		FID ↓	KID ↓	TVD ↓	ρ ↑	FID ↓	KID ↓	TVD ↓	ρ ↑	FID ↓	KID ↓	TVD ↓	ρ ↑
$\epsilon = 2$	DP-TabImage	241.8	0.326	0.083	-0.012	203.4	0.268	0.155	0.016	259.8	0.237	0.219	0.047
	DP-ImageTab	233.2	0.328	0.439	-0.013	203.6	0.289	0.438	-0.009	318.2	0.314	0.742	-0.012
	DP-JointDM	301.5	0.440	0.507	0.013	295.1	0.384	0.454	0.018	280.0	0.242	0.702	-0.004
	DP-Indep	195.6	0.272	0.092	-0.009	211.9	0.303	0.139	0.003	296.8	0.293	0.216	0.000
	DP-JointMERF	502.5	0.783	0.445	–	438.9	0.687	0.445	–	389.1	0.466	0.590	–
	DPLongSyn	**94.3**	**0.090**	**0.077**	**0.044**	**146.5**	**0.199**	**0.119**	**0.063**	**116.7**	**0.050**	**0.185**	**0.146**
$\epsilon = 8$	DP-TabImage	164.1	0.213	0.052	0.036	184.6	0.248	0.063	-0.006	225.6	0.203	0.196	0.041
	DP-ImageTab	183.7	0.251	0.456	-0.012	210.1	0.306	0.428	0.010	284.7	0.269	0.742	0.019
	DP-JointDM	278.8	0.401	0.459	0.023	276.2	0.398	0.446	0.021	349.5	0.378	0.739	0.038
	DP-Indep	137.7	0.178	0.046	-0.005	184.8	0.262	0.063	0.006	244.5	0.226	0.191	-0.005
	DP-JointMERF	502.5	0.782	0.445	–	438.9	0.687	0.445	–	389.0	0.464	0.585	–
	DPLongSyn	**93.8**	**0.099**	**0.029**	**0.161**	**153.9**	**0.209**	**0.057**	**0.073**	**116.1**	**0.052**	**0.125**	**0.169**
$\epsilon = 32$	DP-TabImage	116.9	0.144	0.047	0.026	**152.1**	0.207	0.049	0.017	198.4	0.170	0.188	0.030
	DP-ImageTab	128.4	0.168	0.465	-0.018	191.3	0.281	0.436	0.020	218.8	0.191	0.735	0.023
	DP-JointDM	272.5	0.387	0.458	0.003	259.2	0.368	0.437	0.015	346.2	0.352	0.746	0.000
	DP-Indep	104.9	0.126	0.052	-0.014	155.8	0.216	0.056	0.023	207.6	0.176	0.183	0.000
	DP-JointMERF	502.5	0.783	0.445	–	438.9	0.686	0.443	–	389.0	0.466	0.587	–
	DPLongSyn	**93.8**	**0.099**	**0.021**	**0.111**	152.3	**0.206**	**0.027**	**0.068**	**113.5**	**0.046**	**0.104**	**0.184**

A notable finding is that the image fidelity of DPLongSyn remains relatively stable across privacy levels. Its FID changes only from 94.3 to 93.8 on ADNI and from 116.7 to 113.5 on CelebV-HQ as ϵ increases from 2 to 32, while most baselines show substantially worse image fidelity under stronger privacy constraints. This is consistent with the design of DPLongSyn: public images provide the source appearance, while private learning focuses on the longitudinal transition. Therefore, privacy noise has a more limited effect on the overall visual appearance of the generated images.

We also find that better image fidelity does not necessarily translate into better longitudinal consistency. For several baselines, FID and KID improve substantially with a larger privacy budget, but the correlation between the recorded and generated changes remains close to zero. This shows that generating realistic individual images is not sufficient to preserve how two observations of the same synthetic subject should differ. For example, on CelebV-HQ, DPLongSyn achieves correlations of 0.146–0.184 across privacy levels, while all baselines remain below 0.05. These results support the central design of DPLongSyn: longitudinal synthesis benefits from modeling the transition directly, instead of treating the two observations as a standard paired image-generation problem.

5.2 Downstream Utility

In this section, we evaluate whether the released synthetic data supports downstream learning on both individual follow-up observations and longitudinal changes.

Follow-up Classification. As shown in Table 2, DPLongSyn achieves the strongest overall downstream performance for follow-up classification. At $\epsilon = 8$, for example, DPLongSyn improves AUC from 0.537 to 0.561 on ADNI, from 0.520 to 0.540 on OASIS-3, and from 0.503 to 0.608 on CelebV-HQ compared with the strongest baseline in each setting.

The results also reveal two failure modes of existing DP synthesis methods. First, some methods fail to preserve the label distribution under privacy constraints. On OASIS-3 at $\epsilon = 2$, DP-ImageTab collapses to a single diagnosis class, while DP-JointDM and DP-Indep also lose classes. DPLongSyn instead preserves all three diagnosis classes across privacy levels, with class proportions close to those of the real data. This agrees with its lower record TVD in Table 1 and provides a more informative training distribution for downstream classifiers.

Second, preserving the label marginal alone is not enough. On CelebV-HQ, several baselines maintain reasonable label distributions but still achieve much lower classification AUC. As shown by the record–image correlation in Table 1, their generated images often fail to express the changes spec-

Table 2: Follow-up classification utility of DP synthetic data. We report macro AUC and balanced accuracy (BAcc), averaged across 24 classifiers.

	Method	ADNI		OASIS-3		CelebV-HQ	
		AUC ↑	BAcc ↑	AUC ↑	BAcc ↑	AUC ↑	BAcc ↑
	Real private data	0.675	0.474	0.632	0.338	0.657	0.184
$\epsilon = 2$	DP-TabImage	0.513	0.347	0.510	0.341	0.554	**0.185**
	DP-ImageTab	0.491	0.313	0.490†	0.338†	0.513	0.161
	DP-JointDM	0.487	0.340	0.473	0.335	0.544	0.172
	DP-Indep	0.514	0.340	0.480	0.324	0.532	0.176
	DP-JointMERF	0.523	0.346	0.497	0.336	0.492	0.141
	DPLongSyn	**0.577**	**0.372**	**0.530**	**0.361**	**0.569**	0.160
$\epsilon = 8$	DP-TabImage	0.537	0.353	0.501	0.324	0.498	0.142
	DP-ImageTab	0.517	0.339	0.498	0.337	0.498	0.165
	DP-JointDM	0.492	0.333	0.520	0.337	0.485	0.158
	DP-Indep	0.503	0.324	0.491	0.335	0.496	0.148
	DP-JointMERF	0.497	0.338	0.500	0.337	0.503	0.161
	DPLongSyn	**0.561**	**0.356**	**0.540**	**0.351**	**0.608**	**0.190**
$\epsilon = 32$	DP-TabImage	0.550	0.360	0.539	0.344	0.501	0.158
	DP-ImageTab	0.504	0.337	0.499	0.331	0.497	0.161
	DP-JointDM	0.467	0.314	0.497	0.335	0.528	0.155
	DP-Indep	0.498	0.329	0.492	0.335	0.487	0.134
	DP-JointMERF	0.498	0.338	0.492	0.337	0.513	0.167
	DPLongSyn	**0.565**	**0.366**	**0.586**	**0.381**	**0.585**	**0.187**

† This release contains one diagnosis class; both scores average the 12/24 evaluable classifiers.

ified by the associated records. DPLongSyn better preserves this correspondence and consequently produces synthetic data that transfers more effectively to real-data classification. These results highlight that downstream utility depends on both preserving the record distribution and maintaining its alignment with the generated images.

Progression Classification. Table 3 shows that DPLongSyn achieves the highest mean progression AUC among the evaluated DP synthesis methods at all three privacy levels. At $\epsilon = 2$, for example, DPLongSyn reaches an AUC of 0.579, compared with 0.539 for the strongest baseline. These results provide further evidence that DPLongSyn preserves longitudinal information that is useful for predicting disease progression.

Table 3: Progression classification AUC on ADNI, averaged across 24 classifiers.

Method	$\epsilon = 2$	$\epsilon = 8$	$\epsilon = 32$
DP-TabImage	0.511	0.509	0.542
DP-ImageTab	0.539	0.522	0.508
DP-JointDM	0.504	0.515	0.465
DP-Indep	0.521	0.518	0.472
DP-JointMERF	0.509	0.514	0.500
DPLongSyn	**0.579**	**0.559**	**0.586**

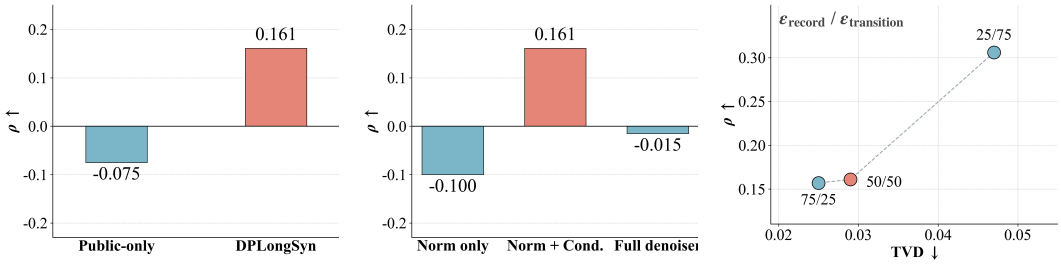
5.3 Empirical Privacy Evaluation

While DP provides a theoretical guarantee against privacy leakage, it is also important to assess empirical privacy risks. In this section, we evaluate DPLongSyn against membership inference attacks (MIAs) (Shokri et al., 2017).

We follow prior work (Hu & Pang, 2023), where the authors instantiate an MIA for diffusion models. The attack is based on the observation that training samples may have lower denoising loss than unseen samples. To apply this attack to DPLongSyn, for each candidate tuple (x^0, a, x^1), we first obtain its transition latent $z = R(x^0, x^1)$ using the registration model. We then compute the denoising loss of z under the latent diffusion model, conditioned on x^0 and a. A lower loss indicates stronger evidence that

Table 4: MIA AUC. $\epsilon \in \{2, 8, 32\}$ denotes DPLongSyn, and $\epsilon = \infty$ denotes non-private training on private data.

Dataset	$\epsilon = 2$	$\epsilon = 8$	$\epsilon = 32$	$\epsilon = \infty$
ADNI	0.494	0.501	0.504	0.613
OASIS-3	0.471	0.472	0.473	0.861
CelebV-HQ	0.455	0.456	0.457	0.788



(a) Private transition adaptation. (b) DP-SGD trainable parameters. (c) Privacy budget allocation.

Figure 3: DPLongSyn ablations on ADNI at $\epsilon = 8$.

the candidate was used for training. We treat private training pairs as members and held-out pairs as non-members, and report the attack AUC in Table 4.

Our results show that DPLongSyn reduces the attack AUC to near-random levels at $\epsilon \in \{2, 8, 32\}$. In contrast, the non-private reference at $\epsilon = \infty$ leads to higher AUCs of 0.613 on ADNI, 0.861 on OASIS-3, and 0.788 on CelebV-HQ. This shows the effectiveness of DPLongSyn in mitigating membership leakage under the evaluated attack.

5.4 Ablation Studies

Effect of Private Transition Adaptation. We compare DPLongSyn with a variant that uses the publicly pretrained transition model; both variants share the same DP records, selected public sources, and sampling settings (Appendix B.9). As shown in Figure 3a, on ADNI at $\epsilon = 8$, private adaptation improves record–transition consistency, increasing Spearman ρ from -0.075 to 0.161. This is likely due to a distribution shift between the public DLBS transitions and the private ADNI transitions: DLBS mostly contains healthy aging, while ADNI includes disease progression. The private adaptation step adapts the transition model to the private transition distribution under DP, improving agreement between the generated transitions and the associated records.

Parameter-Efficient DP Transition Adaptation. During private adaptation, DPLongSyn applies DP-SGD to only a small subset of parameters in the latent diffusion model. To study this design choice, we compare three variants at $\epsilon = 8$ on ADNI: updating only the normalization layers, updating the normalization layers and the record-conditioning MLP as in DPLongSyn, and updating the entire denoiser. All other components of the pipeline are kept fixed. As shown in Figure 3b, DPLongSyn achieves the highest record–transition correlation, with $\rho = 0.161$, compared with -0.100 for normalization-only adaptation and -0.015 for full-denoiser adaptation. These results support jointly adapting the normalization parameters and record-conditioning MLP for record–transition consistency on ADNI at the evaluated privacy budget.

Privacy Budget Allocation. DPLongSyn uses the privacy budget for two components: DP record synthesis and DP transition adaptation. On ADNI, we allocate 25%, 50%, or 75% of the target $\epsilon = 8$ to record synthesis and recalibrate DP-SGD noise to keep the jointly composed budget at $\epsilon = 8$ (Appendix B.8). As shown in Figure 3c, assigning more budget to transition adaptation improves record–transition consistency, with ρ increasing from 0.157 at 75/25 to 0.306 at 25/75. Conversely, assigning more budget to record synthesis improves record fidelity, reducing TVD from 0.047 to 0.025. These results show a clear trade-off between record fidelity and record–transition consistency when allocating the privacy budget.

6 Conclusion

We introduced **DPLongSyn**, a differentially private framework for longitudinal image synthesis that directly models transitions between observations. DPLongSyn first learns a transition prior from public data. It then uses private data to synthesize DP longitudinal records and adapt the transition model to private transitions. Finally, it generates longitudinal samples by applying the learned transitions to public source images. Across two brain MRI datasets and one facial-expression dataset, DPLongSyn improves synthetic data quality and longitudinal consistency, while achieving stronger overall downstream classification performance than existing DP synthesis methods.

AI USE STATEMENT

AI tools were used only to assist with writing polish and debugging limited portions of code. They were not used to develop research ideas, design experiments, or analyze results. The authors reviewed and verified all content in the submission.

ETHICS STATEMENT

This work studies privacy-preserving synthesis of longitudinal image data and is intended to reduce privacy risks associated with sharing sensitive data. Our experiments use medical imaging datasets, including ADNI and OASIS-3, in accordance with their respective access and data-use requirements. More broadly, this work aims to support the responsible sharing and use of sensitive data while protecting individual privacy.

REPRODUCIBILITY STATEMENT

We have aimed to make our work reproducible by providing detailed descriptions of the method, experimental setup, implementation choices, and evaluation procedures in the paper and appendix. We also include source code, configuration files, and instructions for reproducing the experiments in the supplementary material.

REFERENCES

Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318, 2016.

Borja Balle, Gilles Barthe, Marco Gaboardi, Justin Hsu, and Tetsuya Sato. Hypothesis testing interpretations and renyi differential privacy. In *International Conference on Artificial Intelligence and Statistics*, pp. 2496–2506. PMLR, 2020.

Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*, 2018.

Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of cryptography conference*, pp. 635–658. Springer, 2016.

Dingfan Chen, Tribhuvanesh Orekondy, and Mario Fritz. Gs-wgan: A gradient-sanitized approach for learning differentially private generators. *Advances in Neural Information Processing Systems*, 33:12673–12684, 2020.

Hao Chen, Rui Yin, Yifan Chen, Qi Chen, and Chao Li. Learning patient-specific disease dynamics with latent flow matching for longitudinal imaging generation. In *International Conference on Learning Representations*, volume 2026, pp. 82707–82723, 2026a.

Kai Chen, Josephine Lamp, Somesh Jha, and Tianhao Wang. Differentially private paired table-image multimodal synthesis. *arXiv preprint arXiv:2609.00708*, 2026b.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171–4186, 2019.

Tim Dockhorn, Tianshi Cao, Arash Vahdat, and Karsten Kreis. Differentially private diffusion models. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=ZPpQk7FJXF.

Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Chi Nhan Duong, Khoa Luu, Kha Gia Quach, Nghia Nguyen, Eric Patterson, Tien D Bui, and Ngan Le. Automatic face aging in videos via deep reinforcement learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10005–10014. IEEE, 2019.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer, 2006.

Larissa Fischer, Dana Parker, Samira Maboudian, Corrina Fonseca, Claudia Tato-Fernández, Lucie Annen, Prithvi Arunachalam, Julia R Bacci, Michelle Barboure, Serena Capelli, et al. Longitudinal biomarker studies in human neuroimaging: capturing biological change of alzheimer's pathology. *Alzheimer's research & therapy*, 18(1):13, 2025.

Konstantina Giouroukou, Kostas Marias, Manolis Tsiknakis, and Michail E Klontzas. Rethinking privacy in medical imaging ai: from metadata and pixel-level identification risks to federated learning and synthetic data challenges. *Radiology: Artificial Intelligence*, 8(1):e250273, 2025.

Chen Gong, Kecen Li, Zinan Lin, and Tianhao Wang. Dpimagebench: A unified benchmark for differentially private image synthesis. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security*, pp. 4139–4153, 2025.

Frederik Harder, Kamil Adamczewski, and Mijung Park. Dp-merf: Differentially private mean embeddings with randomfeatures for practical privacy-preserving data generation. In *International conference on artificial intelligence and statistics*, pp. 1819–1827. PMLR, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Hailong Hu and Jun Pang. Membership inference of diffusion models. *arXiv preprint arXiv:2301.09956*, 2023.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 2261–2269. Ieee, 2017.

Boah Kim, Inhwa Han, and Jong Chul Ye. Diffusemorph: Unsupervised deformable image registration using diffusion model. In *European conference on computer vision*, pp. 347–364. Springer, 2022.

Akim Kotelnikov, Dmitry Baranchuk, Ivan Rubachev, and Artem Babenko. Tabddpm: Modelling tabular data with diffusion models. In *International conference on machine learning*, pp. 17564–17579. PMLR, 2023.

Pamela J LaMontagne, Tammie LS Benzinger, John C Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G Vlassenko, et al. Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *medrxiv*, pp. 2019–12, 2019.

Zeqi Li, Ruowei Jiang, and Parham Aarabi. Continuous face aging via self-estimated residual age embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15008–15017, 2021.

Zinan Lin, Sivakanth Gopi, Janardhan Kulkarni, Harsha Nori, and Sergey Yekhanin. Differentially private synthetic data via foundation model apis 1: Images. In *International Conference on Learning Representations*, volume 2024, pp. 35109–35159, 2024.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF international conference on computer vision (ICCV)*, pp. 9992–10002. Ieee, 2021.

Daniel Lundqvist, Anders Flykt, and Arne Öhman. Karolinska directed emotional faces. *Cognition and Emotion*, 1998.

Ryan McKenna, Brett Mullins, Daniel Sheldon, and Gerome Miklau. Aim: an adaptive and iterative mechanism for differentially private synthetic data. *Proceedings of the VLDB Endowment*, 15 (11):2599–2612, 2022.

Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pp. 263–275. IEEE, 2017.

Ilya Mironov, Kunal Talwar, and Li Zhang. R\'enyi differential privacy of the sampled gaussian mechanism. *arXiv preprint arXiv:1908.10530*, 2019.

Chi Nhan Duong, Khoa Luu, Kha Gia Quach, and Tien D Bui. Longitudinal face modeling via temporal deep restricted boltzmann machines. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5772–5780, 2016.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Denise C Park, Joseph P Hennessee, Evan T Smith, Micaela Y Chan, Xi Chen, Marianna Dakanali, Michelle E Farrell, Peiying Liu, Hanzhang Lu, Neil Rofsky, et al. The dallas lifespan brain study: A comprehensive adult lifespan data set of brain and cognitive aging. *Scientific Data*, 12(1):846, 2025.

Wonjung Park, Suhyun Ahn, Maria Hernandez, Susana Maniega, and Jinah Park. Brainode: Neural shape dynamics for age-and disease-aware brain trajectories. *Advances in Neural Information Processing Systems*, 38:118550–118581, 2026.

Lemuel Puglisi, Daniel C Alexander, Australian Imaging Biomarkers, Daniele Ravì, Alzheimer's Disease Neuroimaging Initiative, et al. Brain latent progression: Individual-based spatiotemporal disease progression on 3d brain mris via latent diffusion. *Medical image analysis*, pp. 103734, 2025.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 10674–10685. ieee, 2022.

Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)*, pp. 3–18. IEEE, 2017.

Charles Spearman. The proof and measurement of association between two things. 1961.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pp. 6105–6114. PmLR, 2019.

Andrew P Tewkesbury, Alexis J Comber, Nicholas J Tate, Alistair Lamb, and Peter F Fisher. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160:1–14, 2015.

Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International conference on machine learning*, pp. 10347–10357. PMLR, 2021.

Yu-Xiang Wang, Borja Balle, and Shiva Prasad Kasiviswanathan. Subsampled rényi differential privacy and analytical moments accountant. In *The 22nd international conference on artificial intelligence and statistics*, pp. 1226–1235. PMLR, 2019.

Michael W Weiner, Paul S Aisen, Clifford R Jack Jr, William J Jagust, John Q Trojanowski, Leslie Shaw, Andrew J Saykin, John C Morris, Nigel Cairns, Laurel A Beckett, et al. The alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimer's & Dementia*, 6(3): 202–211, 2010.

Yi Feng Wen, Hai Ming Wong, and Colman P McGrath. Longitudinal photogrammetric analysis of soft tissue facial changes: a systematic review of the literature and a summary of recommendations. *Journal of Craniofacial Surgery*, 26(6):1830–1834, 2015.

Nian Wu, Nivetha Jayakumar, Jiarui Xing, and Miaomiao Zhang. Igg: Image generation informed by geodesic dynamics in deformation spaces. In *International Conference on Information Processing in Medical Imaging*, pp. 232–246. Springer, 2025.

Ashkan Yousefpour, Igor Shilov, Alexandre Sablayrolles, Davide Testuggine, Karthik Prasad, Mani Malek, John Nguyen, Sayan Ghosh, Akash Bharadwaj, Jessica Zhao, et al. Opacus: User-friendly differential privacy library in pytorch. *arXiv preprint arXiv:2109.12298*, 2021.

Zhikun Zhang, Tianhao Wang, Ninghui Li, Jean Honorio, Michael Backes, Shibo He, Jiming Chen, and Yang Zhang. {PrivSyn}: Differentially private data synthesis. In *30th USENIX Security Symposium (USENIX Security 21)*, pp. 929–946, 2021.

Hao Zhu, Wayne Wu, Wentao Zhu, Liming Jiang, Siwei Tang, Li Zhang, Ziwei Liu, and Chen Change Loy. Celebv-hq: A large-scale video facial attributes dataset. In *European conference on computer vision*, pp. 650–667. Springer, 2022.

A Additional Method Details

A.1 Public Transition Pretraining

Deformation Representation. We build on the deformation-based generative framework of IGG (Wu et al., 2025) to represent longitudinal transitions. Given a public longitudinal pair $(x^0_{\text{pub}}, a_{\text{pub}}, x^1_{\text{pub}})$, the registration model R encodes the transformation from the source image x^0_{pub} to the follow-up image x^1_{pub} as a latent deformation representation

$$z = R(x^0_{\text{pub}}, x^1_{\text{pub}}).$$

Following IGG, this representation corresponds to a latent velocity trajectory that can be decoded and integrated to obtain a diffeomorphic deformation between the two observations. We adopt the deformation representation and registration procedure of IGG and train R using only the public longitudinal data.

Conditional Transition Modeling. After training R, we encode the public longitudinal pairs into latent transitions and train a conditional latent diffusion model G_{θ_0} to learn

$$z \sim G_{\theta_0}(\cdot \mid x^0_{\text{pub}}, a_{\text{pub}}).$$

Thus, rather than generating the follow-up image directly, G_{θ_0} models a distribution over transformations conditioned on both the source image and its associated longitudinal record. At synthesis time, a sampled latent transition can be decoded through the IGG deformation model and applied to a source image to produce a corresponding follow-up observation.

Record and Source Conditioning. We adapt the conditioning mechanism of IGG to longitudinal image-record data. The record a_{pub} is first mapped to a conditioning representation according to its data type: structured numerical records are encoded directly from their attributes, while records containing categorical semantic descriptions are represented using a text encoder. This conditioning allows the transition model to capture how longitudinal changes depend on the associated record.

The source image is provided as a separate condition so that the generated transition can depend on the anatomy or appearance of the particular source. In our implementation, we use a representation of the full source image rather than the image-gradient condition used in the original IGG formulation. This change preserves more source-specific information and avoids the transition generator becoming effectively insensitive to the source image. The record and source conditions can also be independently dropped during public training to support classifier-free conditional sampling.

All components in this stage, including the registration model R, the transition model G_{θ_0}, and the conditioning transformations, are learned exclusively from $\mathcal{D}_{\text{public}}$. Therefore, public transition pretraining does not access the sensitive dataset and incurs no privacy loss.

A.2 DP Record Synthesis

PrivSyn Overview. PrivSyn (Zhang et al., 2021) is a DP method for synthesizing tabular data from low-order marginal distributions. It captures dependencies among attributes through a set of marginals, privately estimates these marginals, and generates synthetic records that approximately preserve the resulting distributions. In this way, PrivSyn preserves important statistical relationships in the data while providing DP.

PrivSyn for Longitudinal Records. In DPLongSyn, we apply PrivSyn to the record component $\mathcal{A} = \{a_i\}_{i=1}^N$ of the sensitive longitudinal dataset. We define the attribute domains, discretization rules, and marginals in advance using public data or a priori knowledge, without making these choices based on the sensitive data. For brain data, the marginals capture relationships between diagnosis transitions and anatomical changes, while for face data, they capture relationships between target expressions and facial landmark changes. PrivSyn then generates the synthetic longitudinal record dataset

$$\tilde{\mathcal{A}} = \textsc{PrivSyn}(\mathcal{A}),$$

which is later used to condition the longitudinal transition model during synthesis.

A.3 DP TRANSITION ADAPTATION

Private Transition Representation. After public pretraining, we adapt the transition model to the longitudinal patterns in the sensitive dataset. For each private longitudinal pair (x_i^0, a_i, x_i^1), we use the fixed public registration model R to extract its latent transition representation,

$$z_i = R(x_i^0, x_i^1).$$

Since R is trained only on public data and remains fixed during this stage, the private images are used to obtain the latent transitions that serve as training targets for adaptation. This allows the private adaptation stage to operate directly on longitudinal change rather than relearning the image representation itself.

Parameter-Efficient Adaptation. The publicly pretrained model G_{θ_0} provides a general prior over longitudinal transitions, but the transition distribution and its relationship with the associated record may differ between the public and sensitive datasets. We therefore adapt G_{θ_0} using the private tuples (z_i, x_i^0, a_i) with the same conditional diffusion objective used during public pretraining.

For brain MRI, we update the normalization parameters and the numerical record-conditioning MLP. For facial expressions, we keep the BERT encoder fixed and update only normalization parameters. The normalization parameters modulate pretrained feature responses, while adapting the numerical record-conditioning MLP allows the relationship between records and transitions to change. All remaining parameters retain their publicly pretrained values. This restricted adaptation aims to reuse the public transition representation while allowing adjustment to the private population.

DP Optimization. We optimize the selected parameters using DP-SGD (Abadi et al., 2016). At each update, per-example gradients are clipped to bound their sensitivity, and Gaussian noise is added to the aggregated gradient before updating the model parameters. Applying DP-SGD to the private transition tuples produces the adapted transition model G_{θ_1}.

Only the parameters selected for the corresponding dataset are updated during private adaptation, while the remainder of the transition generator stays fixed. The resulting model captures private longitudinal transition patterns under differential privacy and is subsequently used to generate transitions conditioned on synthetic records and public source images. The complete privacy accounting for this adaptation and its composition with DP record synthesis is provided in Appendix A.5.

A.4 LONGITUDINAL DATA SYNTHESIS

For each synthetic record $\tilde{a}_j$, DPLongSyn selects a public source image $\tilde{x}_j^0$ to provide the baseline appearance for the synthetic longitudinal pair. When the public pool does not adequately reflect the baseline characteristics implied by $\tilde{a}_j$, we use record-guided matching to select a more appropriate source; otherwise, the source is sampled directly from the public pool. The public source image provides the baseline appearance, while the transition model determines how that image changes over time.

Conditioned on the selected source image and synthetic record, we sample a latent transition from the DP-adapted model,

$$\tilde{z}_j \sim G_{\theta_1}(\cdot \mid \tilde{x}_j^0, \tilde{a}_j).$$

The sampled latent variable represents the longitudinal transformation to be applied to the source image, rather than a new image generated independently from scratch.

Following the deformation procedure described in Appendix A.1, the sampled transition $\tilde{z}_j$ is decoded into a deformation and applied to $\tilde{x}_j^0$ to obtain the synthetic follow-up image $\tilde{x}_j^1$. Repeating this process produces $\tilde{\mathcal{D}} = \left\{(\tilde{x}_j^0, \tilde{a}_j, \tilde{x}_j^1)\right\}_{j=1}^{M}$.

A.5 PRIVACY GUARANTEE

In this section, we provide the privacy analysis of DPLongSyn.

Theorem A.1 (Privacy of DPLongSyn)**.** *Let the PrivSyn mechanism satisfy ρ_0-zCDP (Zhang et al., 2021), and let the DP-SGD adaptation satisfy $(\alpha, \gamma_1(\alpha))$-RDP for every $\alpha > 1$. Then DPLongSyn*

satisfies (ϵ, δ)-differential privacy, where

$$\epsilon = \min_{\alpha>1} \left\{ \alpha\rho_0 + \gamma_1(\alpha) - \frac{\log\delta + \log\alpha}{\alpha - 1} + \log\left(\frac{\alpha-1}{\alpha}\right) \right\}.$$

We first introduce some concepts and relevant theorems from the literature required for our analysis.

Definition A.2 (Rényi Differential Privacy (Mironov, 2017))**.** For $\alpha > 1$, a randomized mechanism $\mathcal{M}$ satisfies (α, γ)-Rényi differential privacy (RDP) if, for any neighboring datasets $\mathcal{D} \simeq \mathcal{D}'$,

$$D_\alpha\left(\mathcal{M}(\mathcal{D}) \| \mathcal{M}(\mathcal{D}')\right) \leq \gamma,$$

where

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1}\log \mathbb{E}_{x\sim Q}\left[\left(\frac{P(x)}{Q(x)}\right)^\alpha\right].$$

Theorem A.3 (RDP Composition (Mironov, 2017))**.** *If mechanisms $\mathcal{M}_1, \ldots, \mathcal{M}_K$ satisfy $(\alpha, \eta_1), \ldots, (\alpha, \eta_K)$-RDP, respectively, then their composition satisfies*

$$\left(\alpha, \sum_{k=1}^K \eta_k\right)\text{-RDP}.$$

Theorem A.4 (RDP to DP conversion (Balle et al., 2020))**.** *If a mechanism satisfies (α, γ)-RDP, then for any $0 < \delta < 1$, it satisfies (ϵ, δ)-DP with*

$$\epsilon = \gamma - \frac{\log\delta + \log\alpha}{\alpha - 1} + \log\left(\frac{\alpha-1}{\alpha}\right).$$

We also use zero-concentrated differential privacy (zCDP) (Bun & Steinke, 2016) for the PrivSyn mechanism to synthesize DP records.

Definition A.5 (Zero-Concentrated Differential Privacy (Bun & Steinke, 2016))**.** A randomized mechanism $\mathcal{M}$ satisfies ρ-zero-concentrated differential privacy (zCDP) if, for every $\alpha > 1$ and any neighboring datasets $\mathcal{D} \simeq \mathcal{D}'$,

$$D_\alpha\left(\mathcal{M}(\mathcal{D}) \| \mathcal{M}(\mathcal{D}')\right) \leq \alpha\rho.$$

Equivalently, a ρ-zCDP mechanism satisfies $(\alpha, \alpha\rho)$-RDP for every $\alpha > 1$.

Theorem A.6 (zCDP to DP Conversion (Bun & Steinke, 2016))**.** *If a mechanism satisfies ρ-zCDP, then for any $0 < \delta < 1$, it satisfies*

$$\left(\rho + 2\sqrt{\rho\log(1/\delta)}, \delta\right)\text{-DP}.$$

We next introduce the Gaussian mechanism and DP-SGD used in DPLongSyn.

Definition A.7 (Gaussian Mechanism)**.** Let f be a vector-valued query with ℓ_2 sensitivity

$$\Delta_2(f) = \max_{\mathcal{D}\simeq\mathcal{D}'} \|f(\mathcal{D}) - f(\mathcal{D}')\|_2.$$

The Gaussian mechanism releases

$$\mathcal{M}(\mathcal{D}) = f(\mathcal{D}) + \mathcal{N}\left(0, \sigma^2\Delta_2(f)^2 I\right),$$

where σ is the noise multiplier.

Theorem A.8 (RDP of the Gaussian Mechanism (Mironov, 2017))**.** *For any $\alpha > 1$, the Gaussian mechanism with noise multiplier σ satisfies*

$$\left(\alpha, \frac{\alpha}{2\sigma^2}\right)\text{-RDP}.$$

When each record is independently included with probability q before applying the Gaussian mechanism, the resulting mechanism is called the sampled Gaussian mechanism.

Theorem A.9 (RDP of the Sampled Gaussian Mechanism (Mironov et al., 2019; Wang et al., 2019))**.** *For a sampled Gaussian mechanism with Poisson sampling rate q and noise multiplier σ, let $\gamma_{\mathrm{SGM}}(\alpha; q, \sigma)$ denote its order-α RDP cost. Then, for every $\alpha > 1$, the mechanism satisfies*

$$(\alpha, \gamma_{\mathrm{SGM}}(\alpha; q, \sigma))\text{-RDP}.$$

DP-SGD (Abadi et al., 2016) trains a model under DP by clipping per-example gradients and applying Gaussian noise to their aggregate. With Poisson subsampling, each optimization step corresponds to a sampled Gaussian mechanism. Therefore, after T optimization steps, RDP composition gives

$$\gamma_1(\alpha) = T\,\gamma_{\mathrm{SGM}}(\alpha; q, \sigma),$$

and the DP-SGD adaptation satisfies $(\alpha, \gamma_1(\alpha))$-RDP.

We are ready to do the privacy analysis of DPLongSyn (Theorem A.1).

Proof. DPLongSyn accesses the sensitive dataset through two private mechanisms: PrivSyn for DP record synthesis and DP-SGD for transition adaptation.

Following Zhang et al. (2021), PrivSyn satisfies ρ_0-zCDP. By the definition of zCDP, it therefore satisfies

$$(\alpha, \alpha\rho_0)\text{-RDP}$$

for every $\alpha > 1$.

The DP-SGD adaptation satisfies

$$(\alpha, \gamma_1(\alpha))\text{-RDP},$$

where $\gamma_1(\alpha)$ is the RDP cost accumulated over all DP-SGD optimization steps.

Applying Theorem A.3, the composition of PrivSyn and DP-SGD satisfies

$$(\alpha, \alpha\rho_0 + \gamma_1(\alpha))\text{-RDP}.$$

Then, by Theorem A.4, for any $0 < \delta < 1$, the composed mechanism satisfies (ϵ, δ)-DP with

$$\epsilon = \alpha\rho_0 + \gamma_1(\alpha) - \frac{\log\delta + \log\alpha}{\alpha - 1} + \log\left(\frac{\alpha - 1}{\alpha}\right).$$

Optimizing over the Rényi order gives

$$\epsilon = \min_{\alpha > 1}\left\{\alpha\rho_0 + \gamma_1(\alpha) - \frac{\log\delta + \log\alpha}{\alpha - 1} + \log\left(\frac{\alpha - 1}{\alpha}\right)\right\}.$$

After PrivSyn and DP-SGD adaptation, all subsequent stages of DPLongSyn depend only on public data or outputs of these two DP mechanisms. In particular, synthetic record sampling, public source selection, transition sampling, and follow-up image generation depend only on public data and DP outputs. By the post-processing property of DP, these operations incur no additional privacy loss.

Therefore, the final synthetic longitudinal dataset produced by DPLongSyn satisfies (ϵ, δ)-DP. □

B Experimental Supplementary

B.1 Computational Resources

Our experiments were run on four NVIDIA H100 PCIe 80 GB GPUs. Using a single H100 GPU, DPLongSyn takes approximately 1.5 hours to generate 2,000 synthetic brain MRI samples and 2.5 hours to generate 2,000 synthetic facial-expression samples.

B.2 Datasets

We evaluate DPLongSyn on three settings across two domains: brain MRI and facial expressions. Table 5 summarizes the datasets and data splits used in our experiments. All images are represented as 128×128 single-channel images with intensities in $[0, 1]$. Public datasets are divided into training and validation splits, while private test splits are reserved for final evaluation.

Brain MRI datasets. We use the Dallas Lifespan Brain Study (DLBS) (Park et al., 2025) as the public brain dataset. DLBS contains longitudinal T1-weighted MRI scans from cognitively normal adults across the adult lifespan, with up to three imaging waves per subject. We construct forward longitudinal pairs between available visits of the same subject, resulting in 535 training and 137 validation pairs.

Table 5: Datasets used in our experiments after preprocessing and pair construction.

Domain	Dataset	Split	# Pairs
Brain MRI	DLBS	Public train	535
	DLBS	Public validation	137
	ADNI	Private train	1,095
	ADNI	Private test	126
	OASIS-3	Private train	380
	OASIS-3	Private test	164
Faces	KDEF	Public train	542
	KDEF	Public validation	65
	CelebV-HQ	Private train	779
	CelebV-HQ	Private validation	91
	CelebV-HQ	Private test	97

We use ADNI (Weiner et al., 2010) and OASIS-3 (LaMontagne et al., 2019) as two separate private datasets. For each subject, we construct one longitudinal pair using the earliest available scan as the source and the latest available scan as the follow-up. After preprocessing and clinical-data matching, ADNI contains 1,095 training and 126 test pairs, while OASIS-3 contains 380 training and 164 test pairs. All private splits are subject-disjoint.

For all three brain datasets, the MR images undergo skull stripping, bias-field correction, and affine registration to the MNI152 template. We extract the same fixed coronal slice at MNI $y = -20$ mm from each scan. Each slice is then cropped around the brain, resized to 128×128, and independently normalized to $[0, 1]$. For each longitudinal pair, we additionally align the follow-up slice to its source using rigid rotation and translation before estimating the longitudinal deformation.

Each brain record describes the change between the two visits. It contains the diagnosis transition, age at the first visit, time between visits, and changes in hippocampal and ventricular volumes.

Facial-expression datasets. We use KDEF (Lundqvist et al., 1998) as the public face dataset. KDEF contains posed facial expressions captured under controlled conditions. We use frontal-view images and construct neutral-to-expression pairs from the same identity, resulting in 542 training and 65 validation pairs with identity-disjoint splits.

We use CelebV-HQ (Zhu et al., 2022) as the private face dataset. CelebV-HQ contains Internet face videos with expression annotations. For each selected video, we construct a pair using a frame from a neutral segment as the source and a frame near the peak expression as the follow-up. This produces 779 training, 91 validation, and 97 test pairs, with splits made by video ID.

We apply the same face preprocessing to KDEF and CelebV-HQ. Facial landmarks are first detected, and each face is aligned to a shared template using a similarity transform based on the eyes, nose, and mouth corners. The aligned face is then cropped, converted to grayscale, and resized to 128×128. We remove pairs with large head-pose differences so that the observed change primarily reflects facial expression rather than head motion.

Each face record describes the geometric change between the two images. It contains four landmark-based changes: mouth opening, mouth width, brow raise, and eye opening. These measurements are normalized by the inter-ocular distance and stored as differences between the source and follow-up images.

B.3 Evaluation Metric Details

Record fidelity. We compute total variation distance (TVD) separately for each record attribute and report the average across attributes. The reference distribution is computed from the real training records. For brain MRI, we evaluate eight attributes: `dx_from`, `dx_to`, `surv_event`, `age_from`, `age_gap`, `hippocampus_delta`, `ventricles_delta`, and `surv_time`. For facial expressions, we evaluate `expression_to` and the four landmark-change attributes. We exclude `expression_from`, which is neutral for every face pair by construction. Numerical attributes are discretized using the same fixed bins for all methods. Anatomical and facial landmark

changes use bins based on public-data deciles, with outer edges set by the prespecified attribute ranges. We use five-year bins for `age_from`, one-year bins for `age_gap` with a final 10–15-year bin, and fixed yearly bins for `surv_time`. The reported TVD measures marginal record fidelity and does not assess dependencies among attributes.

Image fidelity. We compute FID and KID between 2,000 synthetic follow-up images from each method and the real follow-up images in the held-out test split. We use the 2,048-dimensional Inception-v3 pool-3 features implemented in `torchmetrics` version 1.9.0. Grayscale images of size 128×128 with values in $[0, 1]$ are converted to 8-bit images and replicated across three channels. The feature extractor resizes the images to 299×299. KID uses the unbiased polynomial-kernel MMD estimator, averaged over 100 random subsets. The subset size is 50, capped at half the size of the smaller comparison set.

Longitudinal consistency and measurement success. The brain MRI ventricular-area estimator is specified and validated in Appendix B.7. It returns a measurement for every real and synthetic brain image pair, so no brain pair is excluded because of measurement failure. For facial expressions, we extract 68 facial landmarks using the `face_alignment` FAN model. We compute mouth opening, mouth width, brow raise, and eye opening using the same landmark definitions as the corresponding record attributes, normalized by interocular distance. An image pair enters the correlation calculation only when landmarks are detected in both images. Table 6 reports the fraction of measurable synthetic pairs for each method. We report a correlation as "–" when no valid pairs are available, rather than replacing an undefined correlation with zero.

Table 6: Fraction of synthetic face pairs with detectable landmarks in both images.

Method	$\epsilon = 2$	$\epsilon = 8$	$\epsilon = 32$	$\epsilon = \infty$
DPLongSyn	0.999	0.999	0.999	0.984
DP-TabImage	0.951	0.943	0.977	0.999
DP-ImageTab	0.796	0.863	0.964	0.999
DP-Indep	0.809	0.942	0.969	0.999
DP-JointDM	0.717	0.450	0.659	0.016
DP-JointMERF	0.000	0.000	0.000	0.000

B.4 DPLongSyn Implementation Details

Public pretraining. Following IGG (Wu et al., 2025), we train the registration model on the public longitudinal pairs and then train a conditional latent diffusion model on the resulting deformation representations. The registration model is trained with Adam using a learning rate of 5×10^{-4}, and the diffusion model is trained with Adam using a learning rate of 10^{-4}. For both stages, we select the checkpoint with the best public validation loss.

Conditioning. For brain MRI, we encode the structured numerical record, including age, diagnosis, and anatomical changes, using an MLP. We use numerical conditioning because the public DLBS records contain only CN-to-CN diagnosis transitions, making the numerical attributes more informative for learning longitudinal variation than a text representation. For facial expressions, we use a frozen `bert-base-uncased` encoder (Devlin et al., 2019).

Private adaptation. We perform private adaptation using DP-SGD implemented with Opacus (Yousefpour et al., 2021). For brain MRI, we update the normalization parameters together with the numerical record-conditioning MLP, while for facial expressions we keep the BERT encoder fixed and update only the normalization parameters. We use Adam with a learning rate of 10^{-4}. The clipping norm is 0.052 for brain MRI, set to the median per-sample gradient norm measured on public DLBS pairs, and 0.065 for facial expressions, tuned using only public data and therefore incurring no privacy cost. For each target privacy level, the DP-SGD noise multiplier is calibrated jointly with the PrivSyn mechanism so that the composed mechanism satisfies the target (ϵ, δ) guarantee.

Synthesis. We use ancestral DDPM sampling with a record guidance scale of 2.0. For brain MRI, we select the public source from the 40 age-nearest candidates using an age-adjusted atrophy ranking conditioned on the diagnosis transition in the synthetic record. For facial expressions, public KDEF

source images are sampled uniformly. For each setting and privacy level, we generate $M = 2{,}000$ synthetic samples.

Numerical Privacy Accounting. For the main experiments, the record mechanism is assigned a target budget of $\epsilon/2$, which is converted to its corresponding zCDP parameter ρ_0. For each dataset and target privacy level, we then calibrate the DP-SGD noise multiplier σ using Opacus (Yousefpour et al., 2021) so that the RDP composition of the record mechanism and DP-SGD yields the target ϵ at $\delta = 10^{-5}$. Table 7 reports the parameters used in all main experiments.

Table 7: Numerical privacy accounting for DPLongSyn. We use $\delta = 10^{-5}$ and Poisson sampling. C is the DP-SGD clipping norm, q the sampling rate, T the number of optimization steps, σ the noise multiplier, and ρ_0 the zCDP parameter of the record mechanism.

Setting	ϵ	C	q	T	σ	ρ_0	ϵ_{comp}	$\alpha^\star$
ADNI	2	0.052	0.0500	5540	8.9493	0.0208	2.000	10.2
ADNI	8	0.052	0.0500	5540	2.8307	0.2977	8.000	3.8
ADNI	32	0.052	0.0500	5540	1.1665	3.4307	31.999	2.0
OASIS-3	2	0.052	0.0526	5263	9.1803	0.0208	1.999	10.2
OASIS-3	8	0.052	0.0526	5263	2.8998	0.2977	7.995	3.8
OASIS-3	32	0.052	0.0526	5263	1.1880	3.4307	31.980	2.0
CelebV-HQ	2	0.065	0.0500	5540	8.9493	0.0208	2.000	10.2
CelebV-HQ	8	0.065	0.0500	5540	2.8307	0.2977	8.000	3.8
CelebV-HQ	32	0.065	0.0500	5540	1.1665	3.4307	31.999	2.0

B.5 Baseline Implementation Details

Baseline methods. We compare against five methods from the DP paired table–image synthesis framework of Chen et al. (2026b). DP-TabImage first synthesizes the tabular record and then generates an image conditioned on the synthetic record. DP-ImageTab reverses this order and synthesizes the record conditioned on the generated image. DP-JointDM jointly models the image and record with a single diffusion model. DP-Indep synthesizes the two modalities independently. DP-JointMERF jointly synthesizes the image and record using a DP random-feature mean embedding.

Since these methods are designed for paired table–image synthesis rather than longitudinal images, we represent each source–follow-up pair as a single multi-channel image and train the baselines to synthesize this paired image jointly with its longitudinal record. All generated source and follow-up images are evaluated separately after synthesis.

Fair comparison. We use the authors' released implementations and published hyperparameters whenever feasible. Each baseline is first pretrained on the same public cohort used by DPLongSyn: DLBS for the brain MRI settings and KDEF for the facial-expression setting. Public pretraining uses each method's own training procedure with privacy noise disabled.

All methods are evaluated at $\epsilon \in \{2, 8, 32\}$ with $\delta = 10^{-5}$. Each baseline retains its published privacy-budget allocation across its private mechanisms, and the remaining DP-SGD noise is calibrated using the method's own privacy accountant to satisfy the target privacy level. All methods generate $M = 2{,}000$ synthetic samples and are evaluated with the same downstream protocol as DPLongSyn.

Adaptations to our datasets. The original baselines were designed for lower-resolution images and larger datasets. We adapt their image models to 128×128 resolution and use a DP-SGD sampling rate of 0.5 with 300 optimization steps for our smaller datasets. For DP-TabImage, we omit the private prototype stage at finite privacy budgets and allocate its privacy budget to the subsequent DP-SGD stage. We made this change to improve the baseline's performance in our setting. On ADNI, the adapted protocol achieves much lower FID than the original published protocol across all three privacy levels. For DP-JointDM on facial expressions, we reduce the learning rate from 3×10^{-4} to 10^{-4} because the original setting diverged during public pretraining. All other settings follow the authors' released configurations.

B.6 DOWNSTREAM EVALUATION

For each downstream task, classifiers are trained using only the released synthetic data and evaluated on the held-out private test set. We use the same classifier configurations for all synthetic releases and do not select hyperparameters using the private test set.

Follow-up Classification. For each synthetic release, we train classifiers using only the released synthetic data and evaluate them on the held-out private test set. For ADNI and OASIS-3, the input is the follow-up brain MRI and the label is the follow-up diagnosis among CN, MCI, and Dementia. For CelebV-HQ, the input is the expressive face image and the label is the corresponding facial expression.

We evaluate each synthetic release using the same set of 24 classifiers. Eighteen classifiers are constructed from three frozen visual backbones and six prediction heads. The frozen backbones are DINOv2 ViT-B/14 (Oquab et al., 2023), ImageNet-pretrained ResNet-50 (He et al., 2016), and ImageNet-pretrained Swin-T (Liu et al., 2021). For each backbone, we train logistic regression, linear SVM, random forest, XGBoost, k-nearest neighbors, and an MLP on the extracted features. We additionally evaluate six end-to-end models initialized from ImageNet-pretrained weights: ResNet-18 (He et al., 2016), DenseNet-121 (Huang et al., 2017), EfficientNet-B0 (Tan & Le, 2019), ViT-S/16 (Dosovitskiy, 2020), DeiT-S/16 (Touvron et al., 2021), and Swin-T (Liu et al., 2021). The end-to-end models are trained for 15 epochs with AdamW using a learning rate of 3×10^{-4}, weight decay of 0.01, and class-balanced cross-entropy. We use the same classifier configurations for all synthetic releases and do not select hyperparameters using the private test set.

We report macro AUC and balanced accuracy. Macro AUC computes ROC AUC for each class present in the test set and averages across classes. Balanced accuracy is the mean recall across classes. For each synthetic release, we average performance over classifiers that produce valid results and report the number evaluated when it is below 24. On OASIS-3 at $\epsilon = 2$, the DP-ImageTab release contains only one follow-up diagnosis class. Logistic regression, linear SVM, XGBoost, and MLP therefore fail for each of the three frozen-feature backbones, leaving 12 of 24 classifiers evaluable. The reported AUC of 0.490 and BAcc of 0.338 average only these 12 classifiers and are marked as incomplete results in Table 2. Rows with missing labels are removed before training rather than treated as an additional class.

Longitudinal Progression Classification. We evaluate progression classification on ADNI. Among subjects with MCI at the baseline visit, the task is to predict whether the subject progresses to Dementia at the follow-up visit. The held-out private test subset contains 70 longitudinal pairs, including 23 positive cases.

We use the same 24 classifier architectures as in follow-up classification, adapted to take both visits as input. For the 18 frozen-feature classifiers, we extract features from the baseline and follow-up images using each frozen backbone and form the pair representation

$$[F(x^0),\ F(x^1),\ F(x^1) - F(x^0)].$$

The resulting representation is then used with the same six prediction heads described above.

For the six end-to-end models, we use a shared backbone to encode the baseline and follow-up images. The two image features and their difference are concatenated and passed to a classification head. The same six architectures are used as in follow-up classification and are trained for 15 epochs with AdamW, a learning rate of 3×10^{-4}, weight decay of 0.01, and class-balanced cross-entropy. We report the mean AUC across the 24 classifiers.

Per-Classifier Results. Table 2 reports follow-up classification performance averaged across the evaluable classifiers. We provide the corresponding results for each individual classifier in Tables 8-19. An entry marked "–" indicates that the classifier could not be trained because the corresponding synthetic release did not contain enough label classes.

Table 3 reports progression classification performance averaged across the 24 classifiers. We additionally report the corresponding results for each individual classifier in Tables 20-23.

Table 8: Per-classifier downstream classification results on ADNI using frozen DINOv2 ViT-B/14 features. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP
$\epsilon = 2$	DP-TabImage	0.441/0.367	0.421/0.367	0.530/0.333	0.486/0.333	0.497/0.333	0.399/0.325
	DP-ImageTab	0.525/0.344	0.532/0.343	0.539/0.333	0.525/0.333	0.484/0.333	0.497/0.310
	DP-JointDM	0.475/0.340	0.488/0.340	0.454/0.333	0.476/0.333	0.401/0.333	0.498/0.356
	DP-Indep	0.479/0.344	0.471/0.359	0.519/0.346	0.531/0.378	0.597/0.386	0.518/0.332
	DP-JointMERF	0.528/0.333	0.537/0.340	0.601/0.394	0.505/0.349	0.538/0.336	0.570/0.346
	DPLongSyn	0.605/0.428	0.610/0.379	0.583/0.372	0.593/0.452	0.563/0.354	0.623/0.439
$\epsilon = 8$	DP-TabImage	0.534/0.379	0.526/0.347	0.496/0.333	0.527/0.340	0.494/0.345	0.503/0.340
	DP-ImageTab	0.462/0.265	0.450/0.283	0.498/0.333	0.546/0.333	0.509/0.339	0.438/0.346
	DP-JointDM	0.478/0.296	0.503/0.339	0.448/0.333	0.448/0.329	0.470/0.324	0.468/0.309
	DP-Indep	0.418/0.277	0.440/0.269	0.478/0.333	0.486/0.353	0.526/0.362	0.411/0.279
	DP-JointMERF	0.528/0.333	0.537/0.340	0.501/0.327	0.357/0.303	0.542/0.352	0.566/0.324
	DPLongSyn	0.601/0.380	0.600/0.406	0.573/0.319	0.543/0.334	0.524/0.326	0.551/0.368
$\epsilon = 32$	DP-TabImage	0.520/0.376	0.546/0.376	0.538/0.324	0.567/0.382	0.523/0.322	0.566/0.416
	DP-ImageTab	0.421/0.315	0.384/0.290	0.575/0.392	0.482/0.295	0.617/0.388	0.447/0.318
	DP-JointDM	0.509/0.389	0.508/0.352	0.467/0.271	0.553/0.348	0.418/0.333	0.391/0.313
	DP-Indep	0.491/0.353	0.496/0.336	0.563/0.352	0.446/0.238	0.526/0.308	0.475/0.300
	DP-JointMERF	0.534/0.333	0.547/0.340	0.535/0.333	0.526/0.369	0.534/0.336	0.559/0.354
	DPLongSyn	0.613/0.383	0.638/0.397	0.623/0.377	0.569/0.333	0.495/0.340	0.516/0.341
$\epsilon = \infty$	DP-TabImage	0.581/0.351	0.618/0.366	0.697/0.442	0.649/0.471	0.604/0.387	0.647/0.462
	DP-ImageTab	0.563/0.359	0.569/0.379	0.647/0.347	0.610/0.344	0.560/0.353	0.571/0.396
	DP-JointDM	0.531/0.340	0.537/0.333	0.600/0.459	0.610/0.359	0.619/0.387	0.515/0.333
	DP-Indep	0.436/0.332	0.430/0.335	0.442/0.267	0.456/0.323	0.409/0.238	0.382/0.200
	DP-JointMERF	0.531/0.333	0.545/0.347	0.535/0.333	0.437/0.327	0.543/0.345	0.555/0.368
	DPLongSyn	0.571/0.316	0.567/0.311	0.620/0.362	0.596/0.329	0.548/0.379	0.535/0.331

Table 9: Per-classifier downstream classification results on ADNI using frozen ImageNet ResNet-50 features. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP
$\epsilon = 2$	DP-TabImage	0.513/0.405	0.490/0.380	0.534/0.333	0.509/0.340	0.493/0.340	0.510/0.346
	DP-ImageTab	0.525/0.295	0.512/0.303	0.458/0.325	0.512/0.344	0.562/0.366	0.478/0.313
	DP-JointDM	0.468/0.329	0.471/0.328	0.522/0.333	0.369/0.333	0.551/0.333	0.487/0.356
	DP-Indep	0.514/0.349	0.525/0.372	0.492/0.266	0.537/0.351	0.494/0.339	0.535/0.372
	DP-JointMERF	0.504/0.331	0.504/0.307	0.498/0.313	0.546/0.354	0.501/0.361	0.505/0.340
	DPLongSyn	0.529/0.354	0.527/0.334	0.591/0.365	0.548/0.369	0.494/0.300	0.560/0.390
$\epsilon = 8$	DP-TabImage	0.516/0.336	0.514/0.343	0.544/0.333	0.490/0.340	0.546/0.349	0.563/0.374
	DP-ImageTab	0.486/0.277	0.476/0.297	0.556/0.324	0.509/0.348	0.542/0.348	0.482/0.296
	DP-JointDM	0.534/0.370	0.520/0.349	0.504/0.325	0.467/0.290	0.506/0.336	0.534/0.375
	DP-Indep	0.507/0.345	0.513/0.296	0.568/0.359	0.554/0.357	0.515/0.324	0.508/0.328
	DP-JointMERF	0.495/0.333	0.503/0.306	0.546/0.342	0.494/0.339	0.473/0.368	0.505/0.340
	DPLongSyn	0.572/0.345	0.547/0.374	0.609/0.345	0.555/0.334	0.550/0.356	0.556/0.353
$\epsilon = 32$	DP-TabImage	0.592/0.432	0.595/0.404	0.465/0.316	0.500/0.313	0.494/0.303	0.549/0.367
	DP-ImageTab	0.491/0.341	0.470/0.314	0.531/0.330	0.571/0.324	0.511/0.343	0.517/0.323
	DP-JointDM	0.492/0.263	0.475/0.309	0.450/0.326	0.401/0.285	0.441/0.308	0.495/0.348
	DP-Indep	0.492/0.322	0.484/0.286	0.398/0.279	0.465/0.280	0.487/0.334	0.500/0.388
	DP-JointMERF	0.505/0.334	0.505/0.307	0.479/0.308	0.538/0.360	0.480/0.352	0.505/0.340
	DPLongSyn	0.535/0.348	0.522/0.365	0.600/0.344	0.616/0.398	0.512/0.342	0.582/0.370
$\epsilon = \infty$	DP-TabImage	0.669/0.468	0.633/0.525	0.687/0.438	0.678/0.449	0.569/0.361	0.675/0.516
	DP-ImageTab	0.557/0.371	0.546/0.394	0.679/0.347	0.597/0.402	0.570/0.343	0.559/0.442
	DP-JointDM	0.502/0.300	0.486/0.328	0.513/0.333	0.536/0.333	0.478/0.328	0.507/0.327
	DP-Indep	0.510/0.328	0.483/0.334	0.473/0.312	0.461/0.256	0.451/0.329	0.510/0.340
	DP-JointMERF	0.504/0.341	0.505/0.312	0.477/0.330	0.532/0.353	0.473/0.368	0.505/0.340
	DPLongSyn	0.475/0.283	0.467/0.305	0.567/0.323	0.568/0.331	0.520/0.345	0.511/0.317

B.7 Validation of Longitudinal Consistency Measurement

Our longitudinal consistency metric compares the change measured from an image pair with the change specified in its associated record. For brain MRI, the record contains the change in 3D ventricular volume, measured in mm^3. Because our image synthesis experiments operate on a fixed 2D coronal slice, we measure the lateral-ventricle area in each image and compute the difference

Table 10: Per-classifier downstream classification results on ADNI using frozen ImageNet Swin-T features. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP
$\epsilon = 2$	DP-TabImage	0.520/0.322	0.506/0.295	0.478/0.339	0.543/0.379	0.644/0.333	0.522/0.325
	DP-ImageTab	0.367/0.249	0.393/0.251	0.440/0.325	0.512/0.326	0.422/0.314	0.439/0.293
	DP-JointDM	0.528/0.355	0.530/0.355	0.426/0.333	0.514/0.333	0.503/0.333	0.497/0.377
	DP-Indep	0.484/0.315	0.497/0.321	0.506/0.349	0.547/0.382	0.543/0.345	0.529/0.349
	DP-JointMERF	0.575/0.433	0.536/0.423	0.493/0.327	0.473/0.344	0.504/0.325	0.448/0.348
	DPLongSyn	0.614/0.398	0.594/0.362	0.635/0.327	0.578/0.332	0.557/0.360	0.640/0.429
$\epsilon = 8$	DP-TabImage	0.540/0.389	0.511/0.365	0.576/0.348	0.586/0.363	0.642/0.341	0.525/0.387
	DP-ImageTab	0.533/0.382	0.535/0.384	0.489/0.333	0.598/0.375	0.525/0.327	0.584/0.389
	DP-JointDM	0.508/0.367	0.509/0.365	0.532/0.333	0.462/0.355	0.484/0.320	0.485/0.327
	DP-Indep	0.496/0.297	0.506/0.333	0.506/0.327	0.487/0.329	0.552/0.339	0.592/0.358
	DP-JointMERF	0.575/0.433	0.538/0.422	0.463/0.360	0.430/0.288	0.507/0.325	0.451/0.346
	DPLongSyn	0.532/0.328	0.526/0.319	0.604/0.378	0.565/0.388	0.537/0.377	0.566/0.339
$\epsilon = 32$	DP-TabImage	0.500/0.304	0.498/0.308	0.566/0.362	0.602/0.404	0.452/0.300	0.545/0.371
	DP-ImageTab	0.498/0.329	0.475/0.320	0.476/0.318	0.449/0.294	0.396/0.273	0.482/0.320
	DP-JointDM	0.483/0.331	0.539/0.314	0.441/0.301	0.422/0.237	0.335/0.244	0.471/0.251
	DP-Indep	0.462/0.304	0.437/0.307	0.495/0.336	0.532/0.358	0.430/0.343	0.489/0.341
	DP-JointMERF	0.583/0.433	0.535/0.435	0.426/0.316	0.403/0.256	0.503/0.325	0.448/0.348
	DPLongSyn	0.517/0.352	0.495/0.321	0.637/0.373	0.540/0.366	0.565/0.395	0.496/0.329
$\epsilon = \infty$	DP-TabImage	0.650/0.421	0.676/0.474	0.666/0.415	0.690/0.446	0.624/0.420	0.665/0.456
	DP-ImageTab	0.515/0.360	0.511/0.336	0.636/0.353	0.575/0.355	0.527/0.349	0.541/0.336
	DP-JointDM	0.485/0.327	0.485/0.285	0.550/0.293	0.571/0.347	0.457/0.337	0.520/0.315
	DP-Indep	0.510/0.364	0.525/0.342	0.455/0.329	0.410/0.299	0.441/0.292	0.481/0.327
	DP-JointMERF	0.583/0.425	0.537/0.442	0.484/0.319	0.463/0.336	0.504/0.325	0.453/0.333
	DPLongSyn	0.537/0.390	0.552/0.365	0.588/0.330	0.587/0.391	0.555/0.325	0.581/0.376

Table 11: Per-classifier downstream classification results on ADNI using end-to-end fine-tuned models. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	ResNet-18	DenseNet-121	EffNet-B0	ViT-S/16	DeiT-S/16	Swin-T
$\epsilon = 2$	DP-TabImage	0.455/0.337	0.558/0.333	0.579/0.443	0.567/0.340	0.554/0.349	0.550/0.341
	DP-ImageTab	0.406/0.234	0.543/0.327	0.501/0.296	0.469/0.282	0.612/0.333	0.536/0.337
	DP-JointDM	0.491/0.338	0.463/0.287	0.517/0.370	0.541/0.354	0.530/0.327	0.483/0.340
	DP-Indep	0.451/0.271	0.473/0.329	0.497/0.342	0.519/0.298	0.588/0.333	0.480/0.333
	DP-JointMERF	0.458/0.306	0.500/0.333	0.514/0.352	0.518/0.332	0.541/0.347	0.644/0.333
	DPLongSyn	0.604/0.380	0.605/0.391	0.515/0.360	0.529/0.354	0.570/0.330	0.576/0.359
$\epsilon = 8$	DP-TabImage	0.497/0.322	0.547/0.364	0.493/0.295	0.565/0.380	0.603/0.397	0.551/0.354
	DP-ImageTab	0.463/0.321	0.467/0.305	0.536/0.381	0.602/0.384	0.605/0.439	0.519/0.333
	DP-JointDM	0.458/0.326	0.450/0.294	0.535/0.305	0.465/0.303	0.555/0.365	0.480/0.350
	DP-Indep	0.424/0.248	0.455/0.311	0.549/0.379	0.525/0.311	0.530/0.333	0.529/0.333
	DP-JointMERF	0.408/0.275	0.500/0.333	0.531/0.377	0.447/0.333	0.503/0.273	0.542/0.333
	DPLongSyn	0.577/0.381	0.563/0.417	0.523/0.311	0.544/0.320	0.569/0.329	0.584/0.431
$\epsilon = 32$	DP-TabImage	0.462/0.306	0.597/0.341	0.603/0.422	0.657/0.386	0.652/0.441	0.606/0.370
	DP-ImageTab	0.439/0.283	0.548/0.359	0.585/0.425	0.589/0.340	0.555/0.409	0.578/0.440
	DP-JointDM	0.410/0.356	0.452/0.262	0.505/0.323	0.554/0.372	0.475/0.347	0.527/0.360
	DP-Indep	0.447/0.264	0.503/0.322	0.601/0.448	0.603/0.439	0.567/0.333	0.557/0.333
	DP-JointMERF	0.423/0.339	0.500/0.333	0.516/0.360	0.421/0.226	0.451/0.333	0.494/0.333
	DPLongSyn	0.612/0.413	0.580/0.386	0.527/0.344	0.599/0.374	0.583/0.392	0.590/0.395
$\epsilon = \infty$	DP-TabImage	0.719/0.549	0.681/0.453	0.634/0.449	0.737/0.528	0.713/0.491	0.796/0.492
	DP-ImageTab	0.594/0.405	0.602/0.426	0.569/0.386	0.627/0.449	0.606/0.398	0.683/0.333
	DP-JointDM	0.491/0.337	0.528/0.330	0.507/0.326	0.576/0.397	0.580/0.363	0.598/0.420
	DP-Indep	0.460/0.311	0.513/0.311	0.560/0.392	0.457/0.262	0.467/0.340	0.403/0.333
	DP-JointMERF	0.421/0.324	0.500/0.333	0.502/0.334	0.409/0.333	0.477/0.333	0.384/0.333
	DPLongSyn	0.613/0.420	0.567/0.370	0.539/0.353	0.602/0.351	0.612/0.431	0.589/0.352

between the follow-up and source images. Specifically, we identify low-intensity connected components within a fixed central region and use the source-image brain mask to prevent changes in overall head extent from affecting the measurement. The resulting image-based change is measured in pixels and is therefore not treated as a direct estimate of the recorded 3D volume change. Instead, we

Table 12: Per-classifier downstream classification results on OASIS-3 using frozen DINOv2 ViT-B/14 features. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP
$\epsilon = 2$	DP-TabImage	0.490/0.331	0.471/0.333	0.466/0.333	0.500/0.333	0.496/0.333	0.584/0.333
	DP-ImageTab	–	–	0.500/0.333	–	0.500/0.333	–
	DP-JointDM	0.395/0.333	0.427/0.333	0.446/0.333	0.500/0.333	0.500/0.333	0.465/0.342
	DP-Indep	0.409/0.351	0.405/0.351	0.468/0.333	0.523/0.333	0.449/0.333	0.445/0.333
	DP-JointMERF	0.479/0.333	0.521/0.373	0.478/0.333	0.454/0.333	0.423/0.333	0.560/0.333
	DPLongSyn	0.550/0.383	0.557/0.406	0.572/0.326	0.514/0.310	0.595/0.370	0.570/0.342
$\epsilon = 8$	DP-TabImage	0.512/0.326	0.570/0.319	0.436/0.333	0.455/0.333	0.521/0.333	0.537/0.329
	DP-ImageTab	0.362/0.333	0.399/0.333	0.551/0.333	0.497/0.333	0.500/0.333	0.454/0.333
	DP-JointDM	0.588/0.331	0.576/0.333	0.426/0.333	0.502/0.333	0.500/0.333	0.574/0.333
	DP-Indep	0.509/0.359	0.503/0.314	0.383/0.333	0.535/0.333	0.410/0.333	0.390/0.221
	DP-JointMERF	0.484/0.333	0.547/0.344	0.559/0.333	0.576/0.333	0.423/0.333	0.520/0.333
	DPLongSyn	0.531/0.337	0.512/0.320	0.578/0.364	0.563/0.392	0.588/0.357	0.567/0.381
$\epsilon = 32$	DP-TabImage	0.554/0.391	0.530/0.368	0.438/0.333	0.488/0.333	0.536/0.333	0.499/0.337
	DP-ImageTab	0.570/0.333	0.533/0.329	0.516/0.333	0.543/0.333	0.657/0.333	0.599/0.333
	DP-JointDM	0.510/0.333	0.491/0.333	0.504/0.333	0.451/0.333	0.491/0.333	0.487/0.312
	DP-Indep	0.544/0.319	0.539/0.341	0.551/0.333	0.561/0.333	0.573/0.333	0.506/0.312
	DP-JointMERF	0.455/0.333	0.452/0.239	0.451/0.333	0.469/0.333	0.423/0.333	0.526/0.333
	DPLongSyn	0.503/0.414	0.579/0.328	0.615/0.351	0.621/0.383	0.576/0.329	0.539/0.390
$\epsilon = \infty$	DP-TabImage	0.588/0.367	0.597/0.388	0.701/0.333	0.709/0.322	0.588/0.333	0.572/0.343
	DP-ImageTab	0.580/0.331	0.563/0.333	0.605/0.333	0.552/0.333	0.580/0.333	0.637/0.326
	DP-JointDM	0.596/0.308	0.577/0.353	0.521/0.333	0.631/0.333	0.588/0.333	0.546/0.333
	DP-Indep	0.428/0.295	0.451/0.306	0.532/0.333	0.620/0.333	0.673/0.333	0.563/0.325
	DP-JointMERF	0.496/0.333	0.479/0.333	0.533/0.333	0.632/0.333	0.423/0.333	0.505/0.333
	DPLongSyn	0.611/0.366	0.654/0.508	0.673/0.333	0.685/0.453	0.617/0.333	0.630/0.424

Table 13: Per-classifier downstream classification results on OASIS-3 using frozen ImageNet ResNet-50 features. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP
$\epsilon = 2$	DP-TabImage	0.520/0.489	0.470/0.387	0.615/0.333	0.499/0.333	0.502/0.333	0.518/0.370
	DP-ImageTab	–	–	0.500/0.333	–	0.500/0.333	–
	DP-JointDM	0.452/0.333	0.466/0.333	0.373/0.333	0.500/0.333	0.500/0.333	0.471/0.331
	DP-Indep	0.462/0.322	0.461/0.297	0.454/0.333	0.560/0.333	0.502/0.333	0.466/0.318
	DP-JointMERF	0.490/0.329	0.549/0.423	0.528/0.333	0.493/0.333	0.454/0.333	0.491/0.289
	DPLongSyn	0.421/0.320	0.519/0.282	0.543/0.344	0.491/0.323	0.515/0.383	0.489/0.360
$\epsilon = 8$	DP-TabImage	0.352/0.262	0.440/0.219	0.433/0.333	0.541/0.333	0.512/0.333	0.374/0.312
	DP-ImageTab	0.496/0.333	0.578/0.326	0.483/0.333	0.524/0.333	0.500/0.333	0.464/0.331
	DP-JointDM	0.552/0.326	0.511/0.427	0.549/0.333	0.608/0.333	0.493/0.333	0.554/0.333
	DP-Indep	0.556/0.249	0.581/0.440	0.567/0.333	0.611/0.333	0.525/0.333	0.633/0.310
	DP-JointMERF	0.489/0.329	0.547/0.308	0.559/0.333	0.481/0.333	0.434/0.333	0.518/0.355
	DPLongSyn	0.407/0.326	0.379/0.261	0.549/0.320	0.523/0.346	0.598/0.375	0.414/0.307
$\epsilon = 32$	DP-TabImage	0.492/0.355	0.609/0.389	0.520/0.333	0.541/0.333	0.554/0.333	0.412/0.373
	DP-ImageTab	0.473/0.333	0.435/0.324	0.547/0.333	0.404/0.333	0.496/0.333	0.479/0.333
	DP-JointDM	0.518/0.289	0.501/0.337	0.497/0.333	0.544/0.333	0.521/0.333	0.479/0.434
	DP-Indep	0.529/0.355	0.486/0.299	0.441/0.333	0.456/0.333	0.519/0.333	0.420/0.333
	DP-JointMERF	0.621/0.431	0.451/0.363	0.568/0.333	0.455/0.333	0.508/0.333	0.500/0.306
	DPLongSyn	0.596/0.455	0.664/0.481	0.531/0.322	0.563/0.373	0.640/0.393	0.538/0.347
$\epsilon = \infty$	DP-TabImage	0.614/0.456	0.543/0.381	0.658/0.333	0.738/0.333	0.627/0.331	0.660/0.358
	DP-ImageTab	0.487/0.361	0.462/0.330	0.559/0.333	0.563/0.333	0.628/0.333	0.576/0.333
	DP-JointDM	0.421/0.265	0.487/0.281	0.562/0.333	0.559/0.333	0.513/0.333	0.429/0.291
	DP-Indep	0.587/0.336	0.574/0.318	0.587/0.333	0.664/0.333	0.599/0.333	0.641/0.344
	DP-JointMERF	0.497/0.333	0.493/0.268	0.402/0.333	0.523/0.344	0.446/0.333	0.501/0.335
	DPLongSyn	0.663/0.538	0.666/0.591	0.661/0.333	0.666/0.370	0.575/0.333	0.692/0.450

use Spearman correlation to evaluate whether the image pair preserves the ordering of longitudinal changes specified by the records.

We validate this measurement on real longitudinal pairs before applying it to synthetic data. On the 1,095 real ADNI training pairs, the image-measured ventricular-area change is strongly correlated with the recorded ventricular-volume change, with Spearman $\rho = 0.715$ (95% bootstrap CI

Table 14: Per-classifier downstream classification results on OASIS-3 using frozen ImageNet Swin-T features. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP
$\epsilon = 2$	DP-TabImage	0.544/0.333	0.541/0.340	0.386/0.333	0.560/0.333	0.535/0.333	0.524/0.346
	DP-ImageTab	–	–	0.500/0.333	–	0.500/0.333	–
	DP-JointDM	0.516/0.333	0.456/0.333	0.475/0.333	0.500/0.333	0.500/0.333	0.452/0.336
	DP-Indep	0.509/0.298	0.503/0.272	0.450/0.333	0.506/0.333	0.497/0.333	0.580/0.338
	DP-JointMERF	0.458/0.310	0.451/0.314	0.565/0.333	0.507/0.333	0.560/0.333	0.462/0.333
	DPLongSyn	0.493/0.317	0.502/0.463	0.561/0.376	0.450/0.301	0.616/0.431	0.484/0.307
$\epsilon = 8$	DP-TabImage	0.623/0.351	0.509/0.346	0.364/0.333	0.532/0.333	0.537/0.333	0.566/0.366
	DP-ImageTab	0.577/0.333	0.638/0.333	0.442/0.333	0.462/0.333	0.500/0.333	0.509/0.329
	DP-JointDM	0.524/0.333	0.595/0.333	0.457/0.333	0.425/0.333	0.478/0.333	0.506/0.352
	DP-Indep	0.443/0.296	0.415/0.288	0.384/0.333	0.385/0.333	0.502/0.333	0.611/0.503
	DP-JointMERF	0.453/0.347	0.439/0.349	0.401/0.333	0.399/0.333	0.561/0.333	0.474/0.333
	DPLongSyn	0.542/0.247	0.523/0.338	0.514/0.360	0.569/0.349	0.672/0.404	0.508/0.340
$\epsilon = 32$	DP-TabImage	0.606/0.341	0.512/0.324	0.527/0.333	0.540/0.333	0.536/0.333	0.694/0.334
	DP-ImageTab	0.553/0.333	0.519/0.329	0.547/0.333	0.376/0.333	0.498/0.333	0.578/0.341
	DP-JointDM	0.411/0.336	0.411/0.345	0.479/0.333	0.383/0.333	0.504/0.333	0.525/0.333
	DP-Indep	0.478/0.400	0.432/0.338	0.444/0.333	0.407/0.333	0.490/0.333	0.507/0.303
	DP-JointMERF	0.458/0.329	0.454/0.380	0.444/0.333	0.460/0.333	0.498/0.333	0.466/0.333
	DPLongSyn	0.554/0.317	0.619/0.362	0.593/0.351	0.624/0.436	0.627/0.308	0.506/0.389
$\epsilon = \infty$	DP-TabImage	0.610/0.382	0.579/0.384	0.701/0.333	0.682/0.331	0.588/0.363	0.642/0.308
	DP-ImageTab	0.728/0.428	0.770/0.440	0.634/0.333	0.652/0.333	0.631/0.333	0.670/0.351
	DP-JointDM	0.578/0.382	0.542/0.327	0.475/0.333	0.569/0.333	0.476/0.333	0.574/0.380
	DP-Indep	0.471/0.348	0.509/0.356	0.464/0.333	0.600/0.333	0.452/0.333	0.452/0.314
	DP-JointMERF	0.434/0.405	0.401/0.343	0.587/0.333	0.490/0.333	0.502/0.333	0.453/0.333
	DPLongSyn	0.635/0.370	0.603/0.373	0.646/0.333	0.702/0.326	0.595/0.333	0.640/0.431

Table 15: Per-classifier downstream classification results on OASIS-3 using end-to-end fine-tuned models. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	ResNet-18	DenseNet-121	EffNet-B0	ViT-S/16	DeiT-S/16	Swin-T
$\epsilon = 2$	DP-TabImage	0.427/0.344	0.592/0.317	0.515/0.313	0.519/0.363	0.484/0.333	0.477/0.263
	DP-ImageTab	0.325/0.333	0.512/0.333	0.511/0.394	0.522/0.333	0.622/0.333	0.389/0.333
	DP-JointDM	0.260/0.333	0.489/0.333	0.521/0.362	0.519/0.333	0.501/0.333	0.658/0.333
	DP-Indep	0.494/0.289	0.377/0.325	0.641/0.341	0.395/0.284	0.418/0.333	0.539/0.333
	DP-JointMERF	0.536/0.387	0.500/0.333	0.499/0.318	0.425/0.333	0.531/0.333	0.503/0.333
	DPLongSyn	0.489/0.346	0.481/0.298	0.576/0.498	0.431/0.336	0.669/0.417	0.623/0.427
$\epsilon = 8$	DP-TabImage	0.434/0.310	0.586/0.333	0.419/0.305	0.607/0.333	0.506/0.333	0.654/0.333
	DP-ImageTab	0.494/0.333	0.542/0.333	0.523/0.435	0.519/0.333	0.535/0.333	0.414/0.333
	DP-JointDM	0.418/0.329	0.541/0.333	0.462/0.333	0.532/0.326	0.711/0.333	0.388/0.333
	DP-Indep	0.447/0.379	0.307/0.328	0.498/0.365	0.593/0.333	0.485/0.333	0.506/0.333
	DP-JointMERF	0.513/0.326	0.500/0.333	0.529/0.397	0.509/0.333	0.450/0.333	0.629/0.333
	DPLongSyn	0.619/0.395	0.592/0.410	0.499/0.340	0.532/0.357	0.585/0.385	0.596/0.422
$\epsilon = 32$	DP-TabImage	0.445/0.333	0.542/0.333	0.487/0.313	0.599/0.331	0.657/0.361	0.615/0.370
	DP-ImageTab	0.350/0.333	0.408/0.333	0.378/0.280	0.577/0.333	0.554/0.333	0.381/0.333
	DP-JointDM	0.474/0.333	0.459/0.326	0.615/0.339	0.482/0.329	0.606/0.333	0.574/0.333
	DP-Indep	0.346/0.337	0.585/0.358	0.458/0.326	0.549/0.361	0.389/0.333	0.600/0.333
	DP-JointMERF	0.498/0.289	0.495/0.333	0.542/0.418	0.568/0.333	0.520/0.326	0.517/0.333
	DPLongSyn	0.603/0.408	0.633/0.385	0.547/0.385	0.630/0.399	0.592/0.397	0.579/0.444
$\epsilon = \infty$	DP-TabImage	0.638/0.426	0.496/0.340	0.582/0.387	0.639/0.315	0.701/0.369	0.679/0.358
	DP-ImageTab	0.604/0.378	0.549/0.373	0.499/0.278	0.643/0.333	0.573/0.333	0.479/0.333
	DP-JointDM	0.516/0.333	0.583/0.367	0.343/0.203	0.543/0.371	0.567/0.333	0.403/0.333
	DP-Indep	0.690/0.373	0.484/0.314	0.451/0.271	0.587/0.380	0.605/0.333	0.452/0.333
	DP-JointMERF	0.545/0.337	0.500/0.333	0.527/0.360	0.610/0.333	0.529/0.333	0.489/0.333
	DPLongSyn	0.578/0.353	0.443/0.200	0.603/0.487	0.599/0.509	0.683/0.411	0.603/0.333

$[0.677, 0.748]$; 2,000 bootstrap resamples). We observe a similarly strong correlation of $\rho = 0.744$ on the real OASIS-3 training pairs. These results provide a real-data reference for interpreting the substantially lower correlations obtained by synthetic methods and show that the image-based measurement captures meaningful longitudinal variation despite being computed from a single 2D slice.

Table 16: Per-classifier downstream classification results on CelebV-HQ using frozen DINOv2 ViT-B/14 features. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP
$\epsilon = 2$	DP-TabImage	0.541/0.161	0.525/0.174	0.421/0.167	0.629/0.167	0.514/0.167	0.614/0.273
	DP-ImageTab	0.526/0.190	0.610/0.168	0.465/0.167	0.479/0.167	0.525/0.167	0.484/0.169
	DP-JointDM	0.559/0.121	0.540/0.190	0.607/0.167	0.551/0.167	0.450/0.167	0.488/0.099
	DP-Indep	0.533/0.251	0.464/0.213	0.544/0.167	0.526/0.167	0.463/0.167	0.582/0.300
	DP-JointMERF	0.585/0.119	0.496/0.235	0.453/0.167	0.404/0.167	0.467/0.167	0.580/0.111
	DPLongSyn	0.572/0.161	0.569/0.130	0.543/0.167	0.618/0.167	0.670/0.167	0.632/0.203
$\epsilon = 8$	DP-TabImage	0.491/0.113	0.465/0.120	0.525/0.167	0.588/0.167	0.621/0.167	0.551/0.156
	DP-ImageTab	0.499/0.172	0.457/0.194	0.428/0.167	0.457/0.167	0.513/0.167	0.560/0.169
	DP-JointDM	0.465/0.151	0.471/0.159	0.431/0.167	0.558/0.164	0.480/0.164	0.493/0.123
	DP-Indep	0.524/0.057	0.550/0.057	0.535/0.167	0.524/0.167	0.499/0.167	0.645/0.109
	DP-JointMERF	0.584/0.141	0.553/0.132	0.411/0.167	0.561/0.167	0.479/0.167	0.614/0.155
	DPLongSyn	0.619/0.196	0.579/0.198	0.631/0.167	0.661/0.204	0.659/0.208	0.700/0.223
$\epsilon = 32$	DP-TabImage	0.457/0.117	0.559/0.120	0.444/0.167	0.614/0.167	0.537/0.167	0.486/0.152
	DP-ImageTab	0.459/0.179	0.526/0.181	0.469/0.167	0.567/0.167	0.492/0.167	0.551/0.178
	DP-JointDM	0.561/0.167	0.516/0.142	0.459/0.167	0.355/0.167	0.504/0.167	0.549/0.111
	DP-Indep	0.423/0.130	0.411/0.134	0.532/0.167	0.555/0.167	0.525/0.167	0.484/0.071
	DP-JointMERF	0.601/0.206	0.508/0.162	0.523/0.167	0.322/0.167	0.474/0.164	0.593/0.173
	DPLongSyn	0.653/0.380	0.663/0.158	0.558/0.167	0.665/0.160	0.604/0.167	0.692/0.254
$\epsilon = \infty$	DP-TabImage	0.614/0.241	0.594/0.194	0.699/0.167	0.731/0.167	0.663/0.164	0.718/0.237
	DP-ImageTab	0.468/0.183	0.439/0.128	0.618/0.167	0.621/0.167	0.489/0.167	0.560/0.186
	DP-JointDM	0.468/0.177	0.528/0.196	0.536/0.167	0.579/0.167	0.462/0.167	0.553/0.174
	DP-Indep	0.376/0.149	0.391/0.097	0.405/0.167	0.439/0.167	0.515/0.167	0.406/0.135
	DP-JointMERF	0.585/0.142	0.497/0.095	0.569/0.167	0.422/0.167	0.474/0.164	0.578/0.120
	DPLongSyn	0.548/0.165	0.589/0.154	0.672/0.167	0.674/0.167	0.540/0.167	0.664/0.177

Table 17: Per-classifier downstream classification results on CelebV-HQ using frozen ImageNet ResNet-50 features. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP
$\epsilon = 2$	DP-TabImage	0.612/0.220	0.602/0.226	0.524/0.167	0.613/0.167	0.489/0.167	0.511/0.136
	DP-ImageTab	0.452/0.153	0.473/0.134	0.519/0.167	0.644/0.167	0.509/0.167	0.515/0.133
	DP-JointDM	0.618/0.357	0.618/0.134	0.489/0.167	0.514/0.167	0.557/0.164	0.605/0.146
	DP-Indep	0.521/0.147	0.586/0.156	0.564/0.167	0.447/0.167	0.494/0.167	0.566/0.102
	DP-JointMERF	0.517/0.197	0.601/0.074	0.541/0.167	0.441/0.167	0.491/0.167	0.469/0.094
	DPLongSyn	0.511/0.150	0.435/0.168	0.557/0.167	0.607/0.153	0.566/0.164	0.549/0.182
$\epsilon = 8$	DP-TabImage	0.496/0.111	0.428/0.143	0.541/0.167	0.526/0.167	0.472/0.167	0.520/0.172
	DP-ImageTab	0.463/0.133	0.497/0.189	0.502/0.167	0.522/0.167	0.465/0.167	0.468/0.151
	DP-JointDM	0.532/0.194	0.511/0.168	0.498/0.167	0.491/0.167	0.537/0.167	0.519/0.144
	DP-Indep	0.422/0.133	0.380/0.059	0.426/0.167	0.526/0.167	0.445/0.167	0.454/0.159
	DP-JointMERF	0.583/0.305	0.612/0.092	0.612/0.167	0.408/0.167	0.492/0.167	0.494/0.145
	DPLongSyn	0.702/0.220	0.650/0.413	0.619/0.167	0.669/0.167	0.604/0.160	0.603/0.249
$\epsilon = 32$	DP-TabImage	0.446/0.128	0.518/0.088	0.495/0.167	0.487/0.167	0.504/0.167	0.440/0.162
	DP-ImageTab	0.569/0.147	0.463/0.135	0.526/0.167	0.522/0.167	0.476/0.167	0.553/0.144
	DP-JointDM	0.694/0.147	0.615/0.208	0.609/0.167	0.489/0.167	0.524/0.167	0.631/0.158
	DP-Indep	0.554/0.104	0.532/0.084	0.534/0.167	0.493/0.167	0.450/0.167	0.538/0.109
	DP-JointMERF	0.515/0.197	0.618/0.098	0.616/0.167	0.461/0.167	0.470/0.167	0.476/0.119
	DPLongSyn	0.608/0.374	0.617/0.381	0.493/0.167	0.643/0.164	0.522/0.167	0.611/0.237
$\epsilon = \infty$	DP-TabImage	0.577/0.122	0.547/0.090	0.678/0.167	0.703/0.167	0.618/0.167	0.592/0.153
	DP-ImageTab	0.517/0.177	0.509/0.137	0.514/0.167	0.628/0.167	0.466/0.167	0.433/0.142
	DP-JointDM	0.520/0.124	0.459/0.102	0.577/0.167	0.591/0.167	0.508/0.167	0.432/0.102
	DP-Indep	0.408/0.091	0.473/0.128	0.570/0.167	0.547/0.167	0.403/0.167	0.442/0.142
	DP-JointMERF	0.506/0.180	0.632/0.057	0.686/0.167	0.592/0.167	0.438/0.167	0.477/0.119
	DPLongSyn	0.555/0.122	0.489/0.058	0.609/0.167	0.620/0.167	0.585/0.167	0.584/0.184

B.8 Privacy Accounting for the Budget-Allocation Ablation

For a target budget $(\epsilon, \delta) = (8, 10^{-5})$, we set the record allocation to $\epsilon_{\text{rec}} = s\epsilon$, where $s \in \{0.25, 0.50, 0.75\}$. We obtain the record mechanism's parameter ρ_0 by solving

$$\epsilon_{\text{rec}} = \rho_0 + 2\sqrt{\rho_0 \log(1/\delta)}.$$

Table 18: Per-classifier downstream classification results on CelebV-HQ using frozen ImageNet Swin-T features. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP
$\epsilon = 2$	DP-TabImage	0.659/0.297	0.616/0.369	0.419/0.167	0.630/0.167	0.475/0.167	0.571/0.198
	DP-ImageTab	0.433/0.164	0.459/0.162	0.525/0.167	0.495/0.167	0.493/0.167	0.492/0.144
	DP-JointDM	0.541/0.166	0.563/0.143	0.511/0.167	0.542/0.167	0.600/0.164	0.498/0.267
	DP-Indep	0.657/0.324	0.668/0.138	0.596/0.167	0.568/0.167	0.575/0.167	0.509/0.109
	DP-JointMERF	0.496/0.077	0.455/0.027	0.524/0.167	0.395/0.167	0.595/0.167	0.432/0.139
	DPLongSyn	0.515/0.149	0.521/0.128	0.581/0.167	0.577/0.160	0.479/0.167	0.548/0.095
$\epsilon = 8$	DP-TabImage	0.439/0.117	0.482/0.182	0.371/0.167	0.423/0.167	0.523/0.167	0.449/0.067
	DP-ImageTab	0.515/0.140	0.493/0.133	0.494/0.167	0.521/0.167	0.433/0.167	0.497/0.146
	DP-JointDM	0.511/0.162	0.513/0.162	0.449/0.167	0.457/0.164	0.491/0.167	0.449/0.082
	DP-Indep	0.510/0.215	0.445/0.055	0.496/0.167	0.499/0.167	0.587/0.167	0.615/0.314
	DP-JointMERF	0.428/0.154	0.407/0.130	0.447/0.167	0.428/0.164	0.563/0.167	0.414/0.181
	DPLongSyn	0.639/0.159	0.621/0.198	0.581/0.167	0.650/0.208	0.513/0.167	0.647/0.175
$\epsilon = 32$	DP-TabImage	0.638/0.234	0.663/0.334	0.405/0.167	0.435/0.167	0.413/0.167	0.455/0.227
	DP-ImageTab	0.499/0.199	0.415/0.142	0.506/0.167	0.370/0.167	0.485/0.167	0.396/0.106
	DP-JointDM	0.478/0.153	0.556/0.136	0.414/0.167	0.459/0.167	0.500/0.167	0.577/0.065
	DP-Indep	0.413/0.053	0.488/0.102	0.542/0.167	0.449/0.167	0.468/0.167	0.409/0.114
	DP-JointMERF	0.487/0.154	0.430/0.147	0.582/0.167	0.526/0.162	0.549/0.167	0.444/0.222
	DPLongSyn	0.631/0.121	0.590/0.090	0.452/0.167	0.649/0.158	0.457/0.164	0.648/0.104
$\epsilon = \infty$	DP-TabImage	0.520/0.178	0.392/0.124	0.605/0.167	0.684/0.167	0.567/0.167	0.606/0.148
	DP-ImageTab	0.416/0.181	0.451/0.150	0.472/0.167	0.530/0.167	0.487/0.167	0.455/0.160
	DP-JointDM	0.508/0.163	0.574/0.186	0.485/0.167	0.548/0.167	0.481/0.167	0.563/0.113
	DP-Indep	0.535/0.133	0.569/0.106	0.439/0.167	0.487/0.167	0.471/0.167	0.544/0.098
	DP-JointMERF	0.446/0.154	0.428/0.075	0.477/0.167	0.448/0.167	0.591/0.167	0.441/0.222
	DPLongSyn	0.599/0.299	0.659/0.247	0.495/0.167	0.601/0.167	0.592/0.167	0.618/0.315

Table 19: Per-classifier downstream classification results on CelebV-HQ using end-to-end fine-tuned models. Each entry reports macro AUC / balanced accuracy (BAcc). Higher is better for both metrics.

	Method	ResNet-18	DenseNet-121	EffNet-B0	ViT-S/16	DeiT-S/16	Swin-T
$\epsilon = 2$	DP-TabImage	0.623/0.175	0.559/0.181	0.481/0.118	0.571/0.162	0.475/0.088	0.622/0.170
	DP-ImageTab	0.671/0.167	0.555/0.160	0.439/0.168	0.510/0.167	0.466/0.127	0.564/0.167
	DP-JointDM	0.614/0.239	0.470/0.120	0.504/0.087	0.554/0.167	0.514/0.224	0.555/0.167
	DP-Indep	0.499/0.093	0.549/0.311	0.453/0.100	0.428/0.155	0.531/0.167	0.439/0.167
	DP-JointMERF	0.492/0.124	0.500/0.167	0.481/0.093	0.530/0.167	0.404/0.103	0.454/0.167
	DPLongSyn	0.568/0.144	0.625/0.151	0.502/0.150	0.663/0.207	0.652/0.200	0.598/0.137
$\epsilon = 8$	DP-TabImage	0.605/0.109	0.616/0.133	0.475/0.128	0.421/0.072	0.400/0.123	0.516/0.167
	DP-ImageTab	0.616/0.177	0.468/0.164	0.458/0.197	0.613/0.177	0.548/0.144	0.467/0.167
	DP-JointDM	0.468/0.151	0.435/0.151	0.450/0.149	0.557/0.156	0.392/0.144	0.482/0.195
	DP-Indep	0.668/0.211	0.509/0.140	0.417/0.104	0.451/0.105	0.394/0.167	0.393/0.167
	DP-JointMERF	0.499/0.170	0.500/0.167	0.514/0.165	0.488/0.187	0.437/0.076	0.551/0.167
	DPLongSyn	0.577/0.159	0.545/0.162	0.481/0.108	0.506/0.205	0.509/0.098	0.629/0.177
$\epsilon = 32$	DP-TabImage	0.613/0.168	0.571/0.129	0.511/0.098	0.478/0.002	0.486/0.167	0.359/0.167
	DP-ImageTab	0.565/0.168	0.639/0.167	0.423/0.144	0.441/0.142	0.584/0.167	0.429/0.167
	DP-JointDM	0.446/0.167	0.597/0.160	0.399/0.136	0.642/0.147	0.551/0.167	0.550/0.167
	DP-Indep	0.607/0.100	0.539/0.153	0.415/0.150	0.489/0.074	0.464/0.167	0.382/0.167
	DP-JointMERF	0.465/0.151	0.500/0.167	0.555/0.210	0.512/0.173	0.516/0.167	0.562/0.164
	DPLongSyn	0.564/0.149	0.645/0.188	0.515/0.104	0.524/0.124	0.524/0.137	0.509/0.194
$\epsilon = \infty$	DP-TabImage	0.585/0.144	0.671/0.164	0.535/0.164	0.622/0.179	0.667/0.188	0.618/0.140
	DP-ImageTab	0.454/0.201	0.481/0.164	0.504/0.157	0.451/0.149	0.494/0.167	0.448/0.167
	DP-JointDM	0.508/0.192	0.454/0.113	0.408/0.159	0.538/0.160	0.513/0.204	0.537/0.167
	DP-Indep	0.478/0.149	0.434/0.160	0.536/0.325	0.459/0.147	0.495/0.332	0.499/0.194
	DP-JointMERF	0.472/0.153	0.546/0.143	0.472/0.078	0.440/0.135	0.448/0.167	0.575/0.167
	DPLongSyn	0.423/0.159	0.648/0.156	0.560/0.110	0.505/0.133	0.397/0.159	0.444/0.130

For each ρ_0, we recalibrate the DP-SGD noise multiplier σ so that the joint RDP accountant gives $\epsilon = 8$ at $\delta = 10^{-5}$. We use a sampling probability of 0.05 and $5{,}540$ DP-SGD steps for this ablation. The labels 25/75, 50/50, and 75/25 denote the record allocation s; the remaining percentage is not the transition mechanism's standalone privacy budget. For example, under the 25/75 setting, the

Table 20: Progression classification AUC with frozen DINOv2 ViT-B/14 features on ADNI. Mean denotes the average across the six prediction heads.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP	Mean
$\epsilon = 2$	DP-TabImage	0.387	0.396	0.541	0.544	0.370	0.404	0.440
	DP-ImageTab	0.637	0.648	0.528	0.611	0.610	0.614	0.608
	DP-JointDM	0.427	0.393	0.531	0.558	0.480	0.437	0.471
	DP-Indep	0.443	0.433	0.616	0.661	0.535	0.466	0.526
	DP-JointMERF	0.612	0.604	0.547	0.490	0.512	0.507	0.545
	DPLongSyn	0.580	0.565	0.613	0.651	0.417	0.680	0.584
$\epsilon = 8$	DP-TabImage	0.502	0.529	0.487	0.436	0.629	0.488	0.512
	DP-ImageTab	0.609	0.593	0.588	0.332	0.535	0.604	0.543
	DP-JointDM	0.416	0.397	0.619	0.683	0.530	0.362	0.501
	DP-Indep	0.446	0.463	0.718	0.547	0.555	0.437	0.528
	DP-JointMERF	0.612	0.605	0.401	0.459	0.508	0.482	0.511
	DPLongSyn	0.631	0.636	0.629	0.662	0.646	0.632	0.639
$\epsilon = 32$	DP-TabImage	0.720	0.702	0.505	0.651	0.613	0.797	0.665
	DP-ImageTab	0.375	0.363	0.561	0.519	0.845	0.531	0.532
	DP-JointDM	0.375	0.354	0.601	0.564	0.427	0.331	0.442
	DP-Indep	0.555	0.570	0.538	0.457	0.549	0.521	0.532
	DP-JointMERF	0.614	0.611	0.525	0.385	0.517	0.480	0.522
	DPLongSyn	0.636	0.622	0.754	0.637	0.620	0.607	0.646
$\epsilon = \infty$	DP-TabImage	0.560	0.539	0.727	0.694	0.572	0.579	0.612
	DP-ImageTab	0.621	0.510	0.666	0.743	0.562	0.772	0.645
	DP-JointDM	0.512	0.499	0.511	0.503	0.451	0.412	0.481
	DP-Indep	0.433	0.445	0.521	0.437	0.383	0.411	0.438
	DP-JointMERF	0.612	0.605	0.401	0.459	0.508	0.482	0.511
	DPLongSyn	0.560	0.535	0.631	0.655	0.553	0.547	0.580

transition mechanism's standalone privacy parameter is 7.689, rather than 6. Table 24 summarizes the settings and their jointly composed privacy budgets.

B.9 Controls for Private Transition Adaptation

We compare the privately adapted transition model with its publicly pretrained counterpart using the same DP synthetic records, selected public source images, release seed, guidance scale, and latent-scale calibration procedure. Thus, the public-only condition changes the transition-model checkpoint while retaining the DP record synthesis and public-source selection steps. We also evaluate a zero-deformation control that uses the selected public source image as the follow-up image. Table 25 reports image fidelity, record fidelity, longitudinal consistency, and follow-up classification for these matched controls. Private adaptation improves record–transition correlation on both brain benchmarks at $\epsilon = 8$, whereas it does not consistently improve FID or follow-up classification AUC. The correlation for zero deformation is undefined because its measured image change is constant across pairs.

Table 21: Progression classification AUC with frozen ImageNet-pretrained ResNet-50 features on ADNI. Mean denotes the average across the six prediction heads.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP	Mean
$\epsilon = 2$	DP-TabImage	0.545	0.513	0.545	0.374	0.574	0.463	0.502
	DP-ImageTab	0.498	0.549	0.489	0.435	0.532	0.571	0.512
	DP-JointDM	0.452	0.433	0.459	0.448	0.477	0.456	0.454
	DP-Indep	0.472	0.504	0.535	0.613	0.555	0.512	0.532
	DP-JointMERF	0.528	0.530	0.418	0.769	0.637	0.500	0.564
	DPLongSyn	0.639	0.611	0.622	0.684	0.509	0.658	0.620
$\epsilon = 8$	DP-TabImage	0.424	0.385	0.512	0.394	0.491	0.515	0.453
	DP-ImageTab	0.487	0.502	0.579	0.489	0.500	0.586	0.524
	DP-JointDM	0.459	0.502	0.572	0.602	0.382	0.508	0.504
	DP-Indep	0.474	0.451	0.549	0.571	0.513	0.582	0.523
	DP-JointMERF	0.518	0.523	0.636	0.661	0.636	0.500	0.579
	DPLongSyn	0.592	0.543	0.657	0.680	0.608	0.556	0.606
$\epsilon = 32$	DP-TabImage	0.453	0.438	0.556	0.550	0.463	0.589	0.508
	DP-ImageTab	0.424	0.433	0.712	0.546	0.440	0.449	0.501
	DP-JointDM	0.425	0.337	0.435	0.698	0.545	0.513	0.492
	DP-Indep	0.377	0.367	0.448	0.597	0.424	0.461	0.446
	DP-JointMERF	0.509	0.505	0.577	0.589	0.620	0.500	0.550
	DPLongSyn	0.554	0.587	0.649	0.722	0.516	0.540	0.595
$\epsilon = \infty$	DP-TabImage	0.656	0.620	0.727	0.722	0.639	0.633	0.666
	DP-ImageTab	0.610	0.548	0.744	0.683	0.524	0.620	0.621
	DP-JointDM	0.538	0.519	0.544	0.590	0.461	0.488	0.523
	DP-Indep	0.421	0.468	0.342	0.438	0.354	0.486	0.418
	DP-JointMERF	0.518	0.523	0.636	0.661	0.636	0.500	0.579
	DPLongSyn	0.401	0.402	0.718	0.673	0.491	0.444	0.522

Table 22: Progression classification AUC with frozen ImageNet-pretrained Swin-T features on ADNI. Mean denotes the average across the six prediction heads.

	Method	LogReg	Lin-SVM	RF	XGB	kNN	MLP	Mean
$\epsilon = 2$	DP-TabImage	0.592	0.604	0.560	0.610	0.509	0.536	0.568
	DP-ImageTab	0.584	0.564	0.477	0.510	0.462	0.608	0.534
	DP-JointDM	0.623	0.544	0.550	0.606	0.464	0.489	0.546
	DP-Indep	0.417	0.404	0.447	0.511	0.552	0.552	0.480
	DP-JointMERF	0.489	0.500	0.553	0.419	0.518	0.370	0.475
	DPLongSyn	0.708	0.712	0.642	0.611	0.499	0.542	0.619
$\epsilon = 8$	DP-TabImage	0.557	0.565	0.536	0.496	0.576	0.609	0.556
	DP-ImageTab	0.456	0.438	0.612	0.613	0.502	0.516	0.523
	DP-JointDM	0.640	0.667	0.421	0.459	0.518	0.587	0.549
	DP-Indep	0.544	0.537	0.458	0.517	0.585	0.537	0.530
	DP-JointMERF	0.489	0.499	0.544	0.264	0.506	0.360	0.444
	DPLongSyn	0.512	0.506	0.586	0.522	0.505	0.514	0.524
$\epsilon = 32$	DP-TabImage	0.528	0.536	0.584	0.379	0.453	0.525	0.501
	DP-ImageTab	0.436	0.424	0.512	0.525	0.488	0.513	0.483
	DP-JointDM	0.438	0.549	0.419	0.404	0.445	0.321	0.429
	DP-Indep	0.413	0.418	0.477	0.370	0.475	0.444	0.433
	DP-JointMERF	0.492	0.500	0.430	0.394	0.520	0.349	0.448
	DPLongSyn	0.525	0.488	0.638	0.553	0.579	0.575	0.560
$\epsilon = \infty$	DP-TabImage	0.666	0.641	0.698	0.635	0.633	0.672	0.657
	DP-ImageTab	0.547	0.585	0.714	0.643	0.537	0.482	0.585
	DP-JointDM	0.588	0.528	0.529	0.581	0.690	0.537	0.576
	DP-Indep	0.586	0.605	0.401	0.387	0.457	0.463	0.483
	DP-JointMERF	0.489	0.499	0.544	0.264	0.506	0.360	0.444
	DPLongSyn	0.574	0.582	0.539	0.601	0.634	0.568	0.583

Table 23: Progression classification AUC for end-to-end Siamese models on ADNI. Mean denotes the average across the six architectures.

	Method	ResNet-18	DenseNet-121	EffNet-B0	ViT-S/16	DeiT-S/16	Swin-T	Mean
$\epsilon = 2$	DP-TabImage	0.416	0.434	0.612	0.569	0.512	0.644	0.531
	DP-ImageTab	0.337	0.596	0.622	0.396	0.603	0.463	0.503
	DP-JointDM	0.451	0.650	0.474	0.608	0.519	0.562	0.544
	DP-Indep	0.451	0.525	0.511	0.567	0.563	0.669	0.548
	DP-JointMERF	0.436	0.395	0.489	0.534	0.506	0.348	0.451
	DPLongSyn	0.512	0.383	0.441	0.468	0.570	0.573	0.491
$\epsilon = 8$	DP-TabImage	0.289	0.504	0.389	0.692	0.632	0.586	0.515
	DP-ImageTab	0.394	0.606	0.406	0.509	0.630	0.433	0.496
	DP-JointDM	0.574	0.484	0.468	0.577	0.415	0.527	0.507
	DP-Indep	0.365	0.527	0.387	0.588	0.593	0.484	0.491
	DP-JointMERF	0.442	0.500	0.489	0.482	0.673	0.546	0.522
	DPLongSyn	0.569	0.511	0.335	0.395	0.444	0.552	0.468
$\epsilon = 32$	DP-TabImage	0.432	0.628	0.490	0.535	0.506	0.385	0.496
	DP-ImageTab	0.422	0.543	0.558	0.483	0.632	0.466	0.517
	DP-JointDM	0.497	0.586	0.562	0.537	0.365	0.441	0.498
	DP-Indep	0.563	0.525	0.395	0.390	0.399	0.589	0.477
	DP-JointMERF	0.462	0.484	0.489	0.632	0.348	0.461	0.479
	DPLongSyn	0.569	0.444	0.522	0.459	0.589	0.677	0.543
$\epsilon = \infty$	DP-TabImage	0.584	0.714	0.731	0.743	0.687	0.781	0.707
	DP-ImageTab	0.508	0.744	0.652	0.726	0.685	0.539	0.642
	DP-JointDM	0.433	0.572	0.574	0.355	0.460	0.376	0.461
	DP-Indep	0.430	0.348	0.479	0.567	0.519	0.292	0.439
	DP-JointMERF	0.542	0.500	0.511	0.437	0.408	0.401	0.466
	DPLongSyn	0.549	0.507	0.499	0.486	0.524	0.368	0.489

Table 24: Budget-allocation settings on ADNI at a jointly composed privacy budget of $(8, 10^{-5})$.

Record allocation	ϵ_{rec}	ρ_0	σ	Joint ϵ
25%	2.00	0.0800	2.5697	8.000
50%	4.00	0.2977	2.8307	8.000
75%	6.00	0.6269	3.4748	8.000

Table 25: Matched transition controls at $\epsilon = 8$.

Dataset	Transition	FID ↓	TVD ↓	ρ ↑	AUC ↑
ADNI	Private adaptation	93.8	0.029	0.161	0.561
ADNI	Public-only	88.6	0.029	-0.075	0.575
ADNI	Zero deformation	88.9	0.029	–	0.574
OASIS-3	Private adaptation	153.9	0.057	0.073	0.540
OASIS-3	Public-only	153.6	0.057	0.035	0.534
OASIS-3	Zero deformation	132.3	0.057	–	0.568