Soft Computing Fusion with Applications

www.scfa.reapress.com

Soft. Comput. Fusion. Appl. Vol. 3, No. 3 (2026) 194–207.

Paper Type: Original Article

Client-Personalized Differential Privacy for Federated Subgroup Discovery in Heterogeneous Data

Soheil Fakheri[1,*], Omar Mar Cornelio[2], Du Fu[3]

[1] Department of Computer Engineering and Information Technology, La.C., Islamic Azad University, Iran; fakherisoheil@iau.ac.ir.
[2] Centro de Estudio de Matematica Computacional, Universidad de las Ciencias Inform aticas, La Habana, Cuba; omarmar@uci.cu.
[3] Department of Software Engineering, Sichuan Normal University, China; dufu691@gmail.com.

Citation:

Received: 22 February 2026 *Revised: 05 May 2026* *Accepted: 07 July 2026*	Fakheri, S., Mar Cornelio, O., & Fu, D. (2026). Client-personalized differential privacy for federated subgroup discovery in heterogeneous data. *Soft Computing Fusion with Applications, 3*(3), 194-207.

Abstract

Subgroup discovery returns interpretable descriptions whose outcome rates differ from a population baseline. We present FedPSD, a federated design with client-specific noise budgets and precision-weighted subgroup scoring, together with a reported simulation on UCI Adult and the scikit-learn breast-cancer dataset. Twenty eight-client federations cover ten seeds on each corpus. Personalized budgets have a size-weighted mean of 4; this matches a scalar budget-allocation convention, not each record's protection or a worst-case federation guarantee. Compared with the uniform-budget setting of 4, mean held-out top-10 WRAcc rises from 0.0346 to 0.0503 on Adult and from 0.0488 to 0.0559 on breast cancer; the reported pooled paired Wilcoxon result is W=50, p=0.040, with 16 of 20 blocks favourable. Removing precision weighting lowers Adult WRAcc to 0.0127, whereas a uniform budget of 8 under the same machinery achieves 0.0625. The count channel uses sequentially calibrated noise, but raw-data local proposals and pooled preprocessing are uncharged, so end-to-end personalized differential privacy is not established. The weighting changes the population estimand under heterogeneity. Routing and post-stratification are proposed components that were not evaluated, and the centralized search is a deeper heuristic reference rather than a proven upper bound.

Keywords: Subgroup discovery, Personalized differential privacy, Federated analytics, Heterogeneous data, Privacy-utility trade-off, Weighted aggregation.

1 | Introduction

Subgroup discovery searches interpretable descriptions whose outcome rate differs from a baseline. CN2-SD [1] and subsequent survey work [2] establish algorithmic and application context. Symbolic descriptions can support discussion and audit, but interpretable form alone does not establish causal or governance validity.

Corresponding Author: fakherisoheil@iau.ac.ir
https://doi.org/10.22105/scfa.v3i3.110
Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

Institutional privacy preferences may differ. Federated learning provides a communication architecture [3], [4], while statistical heterogeneity affects local and global models [5]. Differential privacy calibrates releases to record adjacency [6], and individual accounting and personalized guarantees address heterogeneous protection [7], [8]. These concepts motivate subgroup search with a fully specified release ledger. They do not themselves establish fairness [9] or privacy for a data-dependent proposal and preprocessing pipeline.

The gap has three faces. First, uniform-budget federated protocols either over-constrain sites that could afford far more accurate statistics or under-protect conservative sites, depending on whether the common budget tracks the strictest or the average participant. Second, personalised noise changes the statistical geometry of aggregation: releases from low-budget clients are noisier, so naive averaging is inefficient, yet weighting by precision can bias the discovered subgroups toward the populations of generous clients. Third, the adaptive search interacts with both problems, since early noisy evaluations steer which descriptions are refined later. This manuscript addresses all three faces at the level of a complete, testable research design.

The contributions combine a conditional privacy design, an explicit aggregation choice, and a reported proof-of-concept campaign. Their measured and untested components are distinguished below:

I. The problem is adaptive subgroup discovery with client-specific record-protection targets. Every released description, preprocessing statistic, count pair, and observable budget-dependent choice must be covered by its client's ledger.

II. FedPSD, a protocol that combines local beam-search proposal of subgroup descriptions, personalised Laplace calibration of subgroup statistics, and inverse-variance federated aggregation of weighted-relative-accuracy scores.

III. The analysis distinguishes precision weighting of normalized rates from population-weighted pooled WRAcc. Under heterogeneous populations these are different estimands; minimum noise variance alone does not establish population representativeness or minimum WRAcc error.

IV. A budget-ledger and query-routing design in which confirmatory queries are steered toward clients with remaining budget, using only public information and already-released aggregates.

V. The reported campaign uses two public corpora and twenty seed-by-corpus cells to compare noisy federated searches, personalized-budget ablations, a local nonprivate variant, and a centralized depth-three heuristic reference.

The balance of the manuscript proceeds in the standard order. Section 2 reviews subgroup discovery, differential privacy and its personalisation, and federated learning under heterogeneity, closing with a gap synthesis. Section 3 states the research questions, hypotheses, system model, and formal problem. Section 4 presents the FedPSD methodology with its algorithm, privacy accounting, and failure conditions. Section 5 reports the executed experimental campaign, Section 6 debates limitations and alternatives, and Section 7 answers the research questions on the measured evidence.

2 | Related Work

2.1 | Subgroup Discovery and Supervised Descriptive Rule Induction

CN2-SD adapts propositional rule search to subgroup quality [1], while surveys consolidate subgroup and contrast-set discovery [2], [10]. Diversity-oriented selection addresses redundant outputs [11]. WRAcc combines population coverage and outcome-rate deviation, and its count inputs motivate calibrated releases.

Subgroup search is adaptive and combinatorial, while its output is a ranked set of symbolic descriptions. A private federated implementation must calibrate both the selection of descriptions and their numeric evaluations. The reviewed centralized methods motivate the search machinery; an exhaustive absence claim about all private subgroup methods is not established by this review.

2.2 | Differential Privacy, Accounting, and Personalisation

Differential privacy protects a specified unit against changes in the released transcript [6], [12]. Institutions here act as curators of many records, rather than local-DP respondents holding a single record [13]. Composition is needed for every adaptive release [14]. Federated model-training studies use distinct record-level or client-level threat models [15], [16]; their guarantees cannot be transferred to subgroup selection without a new analysis. Membership-inference attacks on trained model outputs [17] offer a related motivation; their results do not measure leakage from the subgroup descriptions and counts released in this campaign.

Personalised differential privacy relaxes the single-budget assumption. Conservative-or-liberal personalisation lets individuals be protected at different strengths while retaining usable semantics [8], and individual privacy accounting tracks per-example losses that differ because of participation patterns [7]. The fairness literature supplies a normative complement: treating participants identically is not always treating them equitably, and awareness of relevant differences can be a design requirement [9]. These threads, however, address fixed estimators or model training. Our contribution is to carry personalisation into an adaptive search process where the query set is endogenous—the analyst does not know in advance which subgroup statistics will be asked—and where the aggregation of heterogeneous-noise releases feeds back into the search trajectory.

2.3 | Federated Learning across Heterogeneous Populations

Federated learning studies coordination and heterogeneity [3], [4], [18], including proximal optimization [19] and matched averaging [20]. Secure aggregation can hide weighted client vectors under stated cryptographic assumptions [21], but a server cannot simultaneously observe each private budget report and have that report hidden by aggregation. This design treats budget choices and agreed auxiliary information as public; protecting them is a separate extension.

2.4 | Gap Synthesis

Table 1 positions the design among the reviewed methods. FedPSD combines client-specific count noise, candidate search, and precision weighting. End-to-end privacy and population-representative scoring remain conditional on the entire release pipeline and the selected estimand.

Table 1. Positioning of FedPSD relative to approach families.

Approach Family	Representative Works	Privacy Regime	Budget Heterogeneity	Output Type
Centralized subgroup discovery	[1], [10], [11]	None	Not applicable	Ranked subgroup descriptions
DP statistics and analytics	[13-15]	Record-level DP	Uniform	Numeric estimates
Personalised DP	[7], [8]	Per-individual / per-client DP	Yes	Mechanisms for fixed queries
Federated learning (models)	[3], [4], [15], [19], [20]	Optional central DP	Rarely considered	Model parameters
FedPSD (this work)	Sections 3-4	Conditional count channel; full pipeline unverified	Heterogeneous nominal budgets	Ranked subgroups with noisy quality

3 | Problem Formulation and Research Questions

3.1 | Research Questions and Hypotheses

The first research question concerns statistical faithfulness under heterogeneous noise: RQ1—when clients calibrate noise to their own declared budgets, how faithfully does the federated subgroup ranking track the ranking a no-privacy federation would produce, and how does this fidelity degrade as the dispersion of budgets across clients grows? Faithfulness is measured by rank correlation and by precision and recall against planted ground truth, since a protocol that returns the right subgroups in the wrong order is still of limited governance value. The question isolates the estimation layer from the search layer by holding the proposal mechanism fixed.

The second research question concerns the allocation of discovery effort: RQ2—can a routing policy that steers confirmatory queries toward clients with remaining budget recover discovery power that uniform-budget operation loses, and what coverage bias does such routing introduce into the discovered set? Routing is attractive because budgets are public and unused budget is wasted budget, yet it risks producing subgroups that describe generous clients' populations disproportionately. The question therefore pairs a power metric with a fairness metric, and treats the pair as a frontier to be traced rather than a point to be claimed.

The third research question concerns population representativeness: RQ3—does precision weighting of heterogeneous-noise releases systematically shift the discovered subgroups toward high-budget clients' populations, and to what extent can a post-stratification correction restore population-representative discovery before variance erodes the ranking? Representativeness matters because subgroup descriptions feed resource-allocation decisions, and a budget-weighted view of the world can misdirect them. Three hypotheses make these questions testable.

H1. Inverse-variance aggregation of personalised releases preserves ranking fidelity better than uniform averaging, with the advantage growing in budget dispersion.

H2. Budget-aware routing increases the number of confirmed subgroups at a fixed total discovery cost, at the price of measurable coverage bias toward high-budget populations.

H3. A post-stratification correction reduces coverage bias at a quantified increase in estimator variance, yielding a representativeness-fidelity frontier that practitioners can traverse with a single tuning parameter.

3.2 | System Model and Notation

Client i holds a table T_i with public, fixed n_i records and binary outcome y. The conditional count mechanism uses replace-one-record adjacency within a client, with fixed public budgets and a fixed or separately privatized description universe. Each pair of subgroup counts has L1 sensitivity at most two. If sizes, bin thresholds, selected features, budgets, or proposed descriptions depend on protected data, their publication or effect on the transcript also requires accounting. The campaign uses such uncharged data-dependent inputs, so the complete E_i record-level guarantee is not established.

Table 2. Notation used throughout the manuscript.

Symbol	Meaning	Domain/Unit
m, i	Number of clients, client index	integer
T_i, n_i, N	Client table, its size, federation size	records
A, y	Attribute schema, binary target	-
S	Subgroup description (conjunction of conditions)	-
n(S), u(S)	Subgroup size, outcome-positive count in S	records
p(S), p_0	Outcome rate in S, overall outcome rate	[0, 1]
q_WRAcc(S)	Weighted relative accuracy of S	[-1, 1]
E_i, eps_i^(t)	Client budget ledger, per-round per-query budget	privacy units
b_i	Laplace noise scale for client i	counts
w_i(t)	Aggregation weight of client i at round t	[0, 1]
B, K	Beam width, output subgroup-set size	integer
R	Number of protocol rounds	integer
rho	Budget dispersion (max-to-min E_i ratio)	>= 1
alpha_i	Post-stratification population weight	[0, 1]

3.3 | Formal Problem Statement

The quality of a subgroup description S under the weighted relative accuracy measure is given in *Eq. (1)*, where n(S) is the number of records satisfying S, p(S) the outcome rate inside S, and p_0 the overall outcome rate. The measure rewards subgroups that combine deviation from the baseline with non-negligible coverage; it is the standard reference quality in the subgroup literature [1], [2].

$$q_{\mathrm{WRAcc}}(S) = \frac{n(S)}{N}\,[p(S) - p^{0}]. \tag{1}$$

For a fixed description S, the released vector contains its size n_i(S) and positive-outcome count u_i(S). Both coordinates have scalar sensitivity one, but the pair has L1 sensitivity at most two. Allocate a pair budget eps_i(t,S) and use independent Laplace noise of scale 2/eps_i(t,S), as in *Eq. (2)*. This gives a conditional pair guarantee under the fixed auxiliary assumptions, before composing other releases.

$$\left(\tilde{n}_i(S), \tilde{u}_i(S)\right) = \left(n_i(S) + \eta_i, u_i(S) + \eta'_i\right), \qquad \eta_i, \eta'_i \sim \mathrm{Laplace}\left(\frac{2}{\varepsilon_i(t,S)}\right). \tag{2}$$

The target is a useful ranked subgroup set under a ledger, communication budget, and specified population estimand. Count-pair budgets sum over candidates and rounds, as in *Eq. (3)*; candidate discovery and preprocessing costs must be included separately. Adaptive candidate choices are post-processing only when they use public information or the preceding private transcript. The ledger for counts alone is not a guarantee for a larger transcript.

$$E_i, \mathrm{pre} + E_i, \mathrm{proposal} + \Sigma_t \Sigma_{S \in Q_t} \varepsilon_i(t,S) + E_i, \mathrm{other} \le E_i. \tag{3}$$

4 | Proposed Methodology: The FedPSD Protocol

4.1 | Architecture Overview

The design uses proposal, calibrated count release, aggregation, and optional routing. Privacy-free internal computation becomes a release when data-dependent descriptions or control choices cross the client boundary.

Proposal expands the description beam internally. A description transmitted because it scores well on private data requires a private selection mechanism and ledger charge, or a public fixed candidate route.

Personalised release: For every candidate in the union beam, each client releases the noisy size and outcome-count pair of *Eq. (2)*, drawing the per-query budget from its ledger according to a declared spending policy.

Aggregation: The coordinator combines releases into federated quality estimates using inverse-variance weights, ranks the beam, and prunes to width B.

Routing: Confirmatory re-queries for the most promising candidates are steered toward clients with remaining budget, and each client's residual ledger is reported for the next round.

4.2 | Local Proposal and the Description Language

Descriptions are conjunctions of agreed categorical conditions or bins, with a maximum depth. A client may compute freely inside its boundary, but exporting the locally preferred description is data-dependent disclosure. The executed local-proposal beam is therefore a simulation component with an unaccounted release, rather than an established private selector [1].

A union beam can combine publicly generated refinements, privately selected local proposals, and previous private outputs. Diversity pruning on an already private transcript is post-processing [11]. The private proposal mechanism and preprocessing contract are prerequisites to an end-to-end guarantee and are not implemented or accounted for in the reported campaign.

4.3 | Personalised Release, Inverse-Variance Aggregation, and Ledgers

Eq. (4) gives inverse-noise-variance weights for normalized client rates when each count has the same scalar sensitivity and per-query budget convention: dividing a Laplace count by n_i gives variance proportional to 1/(n_i eps_i)^2. These weights minimize noise variance only for unbiased estimates of a common rate. Non-IID clients have different rates, so the weighted mixture changes the estimand. Population WRAcc instead requires totals n_hat(S)=sum_i n_tilde_i(S), u_hat(S)=sum_i u_tilde_i(S), public N=sum_i n_i, and a protected or public population baseline. A weighted mixture can be scored, but must not be labelled an unbiased pooled-population estimate or a generally optimal nonlinear WRAcc estimator.

$$w_i(t) = \frac{\left(n_i\, \varepsilon_i(t,S)\right)^2}{\Sigma_j \left(n_j\, \varepsilon_j(t,S)\right)^2}. \tag{4}$$

$$\hat{q}_{\text{population}(S)} = \frac{\hat{u}(S) - \hat{p}^0 \hat{n}(S)}{N}. \tag{5}$$

Budgets are allocated per count pair and composed over every candidate and round. Any adaptive spending policy must be determined by public information or privatized prior outputs and obey a privacy filter [7], [14]. Public per-client budget reports are visible to the coordinator; secure aggregation [21] can instead hide individual preweighted statistic vectors, but does not by itself protect released descriptions, preprocessing, or a data-derived allocation.

Algorithm 1. FedPSD discovery round t.

Input: tables T_i, ledger E_i, beam width B, target y, beam Q_t
1: for all clients i = 1..m concurrently do
2: expand a public candidate beam or use a separately privatized local selector
3: for each candidate S in union beam Q_t do
4: allocate eps_i(t,S), accounting for the whole pair and reserved stages
5: release (n_tilde_i(S), u_tilde_i(S)) via Eq. (2)
6: end for
7: report public residual ledger; privatize data-dependent metadata if required
8: end for
9: server: score the specified population totals or precision mixture; state estimand
10: rank Q_t by q_hat; prune to top-B beam; apply diversity selection
11: route confirmatory re-queries for top candidates to clients with
remaining budget (public information only)
12: update ledgers; broadcast next beam Q_{t+1}
Output: ranked subgroup set with noisy qualities and budget reports

4.4 | Budget-Aware Routing and Representativeness Correction

Routing based on public residual budgets and preceding private outputs can be post-processing as a choice rule. The new confirmatory answers still consume budget. Selecting clients changes population coverage; a minimum-client fairness cap does not establish an unbiased population estimate or a formal fairness guarantee.

Precision weighting favours lower-noise normalized client rates, whereas population weighting uses alpha_i=n_i/N. An interpolated weight w_i(beta)=(1-beta)w_i+beta alpha_i, for beta in [0,1], defines an explicit mixture of estimands, not a generic bias-free post-stratification theorem. Restoring a target population also requires compatible subgroup and outcome totals and a matching baseline. Neither routing nor this interpolation was evaluated.

4.5 | Complexity, Assumptions, and Failure Conditions

Candidate evaluation costs O(n_i |Q_t| c), with additional proposal, selection, and noise work. Communication includes statistics, description identifiers, budget metadata, and any cryptographic overhead; ranking costs at least O(|Q_t| log|Q_t|) without a specialized top-B selector. Zero-size clients are excluded from rate normalization, and an all-zero weight denominator requires a declared fallback. Post-processing may project noisy counts to 0<=u_tilde<=n_tilde<=n_i, but clipping introduces estimator bias and the actual implementation convention must be reported. Record independence is relevant to statistical modelling, not required by the DP definition.

FedPSD has identifiable failure modes, and the evaluation is constructed to surface them rather than conceal them. Each condition below names the mechanism at risk, so that a negative result can be attributed precisely. None of these conditions is hypothetical; each corresponds to a plausible deployment scenario in a multi-site consortium.

Ledger exhaustion: A client that spends its ledger early drops out of confirmations, degrading coverage of its population even if proposal succeeded.

Winner's-curse selection: Ranking on noisy estimates systematically promotes descriptions whose noise realisation was favourable, inflating reported quality; the protocol reports bias diagnostics alongside rankings.

Coverage collapse: Under extreme budget dispersion, routing and precision weighting jointly concentrate discovery on a few generous clients, which the coverage-bias metric is designed to detect.

Schema mismatch: Unaligned binning across clients silently merges distinct subpopulations; the protocol requires an agreed discretisation contract as a precondition.

5 | Experimental Results and Analysis

5.1 | Setup: Data, Federation, and Executed Scale

The source reports UCI Adult subsamples of 5,000 rows with 3,999 train and 1,001 test rows, and the scikit-learn breast-cancer dataset with 569 rows, 454 train and 115 test rows [22], [23]. Reported training baseline rates are 0.259 and 0.628. In the loader, target 1 denotes benign and target 0 malignant; the latter rate therefore cannot be labelled malignant without confirming an explicit recoding. Numeric tertiles and categorical support screening use pooled training data; breast-cancer feature screening uses supervised pooled WRAcc to select six features. These preprocessing operations do not use the held-out test split as described, but their data-dependent information is uncharged for privacy and requires a protected preprocessing stage or independently public inputs.

The source partitions training data into eight clients with a Dirichlet construction, occasionally producing empty clients. Adult's largest client holds 47.6 +/- 5.8% of rows, versus 24.1 +/- 3.4% on breast cancer. The centralized and local beam searches use depth three; federated variants use depth two. Uniform settings 1,4,16

allocate count-noise scale 2 Q_pre/eps, with Q_pre=111 on Adult and 83 on breast cancer. FedPSD uses eps_i=clip(c n_i,1,16), with c chosen to make sum_i n_i eps_i/N=4. This equality is a scalar resource convention; clients and records have unequal protection and the maximum eps_i exceeds 4. Precision weighting and raw-data best local proposals from three highest-budget clients distinguish the evaluated variant. The uncharged proposals, preprocessing and allocation inputs prevent a complete privacy guarantee even if the count-query cap is respected.

The reported Adult allocation places 4.3 +/- 1.5 clients at budget 1 and has a maximum 6.25 +/- 0.41; breast-cancer allocations span 1.17 to 6.02. Results summarize ten seeded resamplings per corpus with held-out top-10 WRAcc, Jaccard against the centralized heuristic search, and transmitted-statistic counts. A centralized beam is not an exhaustive oracle, and different depth and candidate routes affect overlap and quality. Runtime, randomization, actual query caps, zero-client handling, clipping, and the allocation solver require code and per-run outputs for verification. The source reports an implementation in Python with NumPy, SciPy [24], and scikit-learn [23], completed in 2.9 s for the twenty federations; its runtime and computations await the code and per-run results for verification.

5.2|Main Comparison (RQ1)

Table 3 reports mean held-out WRAcc. On Adult, the means are 0.0877 for centralized search, 0.0763 for local-only, 0.0708 for uniform 16, 0.0503 for FedPSD, 0.0346 for uniform 4, and 0.0166 for uniform 1. FedPSD has higher WRAcc than uniform 4 on ten Adult seeds and six breast-cancer seeds, according to the source, with mean relative differences of about 45.4 and 14.5%. The breast-cancer uniform-1 mean is -0.0042; this describes the selected subgroup set under the noisy pipeline, not proof that noise alone causes an outcome. Local proposals, candidate depth, weights, and preprocessing affect the rankings, so attributing a particular description to a noise mechanism requires controlled ablations.

Table 3. Main comparison over ten seeds: Top-10 test WRAcc on the held-out 20% split, Jaccard overlap of the top-10 description set with the centralized top-10, and transmitted statistics (count queries; the centralized row counts pooled raw records instead, and local-only sends 8 x 10 x 2 exact counts).

Method	Adult: Test WRAcc	Adult: Jaccard	bc: test WRAcc	bc: Jaccard	Stats sent (Adult/bc)
Centralized beam (heuristic reference)	0.0877 ± 0.0066	1.000 ± 0.000	0.1178 ± 0.0116	1.000 ± 0.000	3999 / 454
Local-only pooling	0.0763 ± 0.0073	0.374 ± 0.193	0.1133 ± 0.0159	0.620 ± 0.265	160 / 160
Uniform-DP, eps = 1	0.0166 ± 0.0169	0.067 ± 0.060	-0.0042 ± 0.0506	0.132 ± 0.103	1176 / 696
Uniform-DP, eps = 4	0.0346 ± 0.0080	0.095 ± 0.051	0.0488 ± 0.0264	0.158 ± 0.106	1170 / 706
Uniform-DP, eps = 16	0.0708 ± 0.0063	0.309 ± 0.054	0.0870 ± 0.0184	0.274 ± 0.107	1184 / 712
FedPSD (mean-weighted eps = 4)	0.0503 ± 0.0177	0.196 ± 0.084	0.0559 ± 0.0294	0.166 ± 0.122	1245 / 826

The reported Jaccard overlaps with the centralized top ten are 0.196 for FedPSD and 0.095 for uniform 4 on Adult, and 0.166 versus 0.158 on breast cancer. Local-only has overlaps of 0.374 and 0.620. The centralized reference uses deeper descriptions, averaging 2.02 conditions on Adult versus 1.68 for FedPSD and 1.44 for uniform 4. Depth is a possible source of residual differences, but no matched-depth ablation quantifies its contribution.

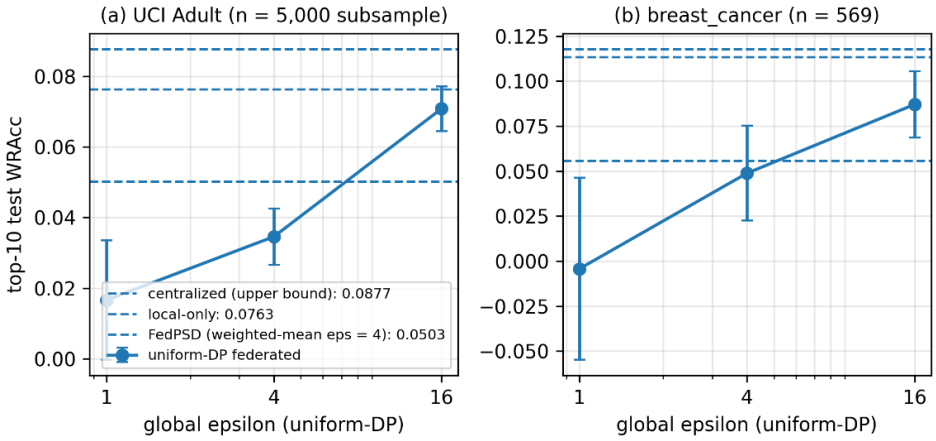

Fig. 1. Reported top-10 test WRAcc versus the nominal uniform count-stage budget, with personalized FedPSD at size-weighted mean 4.

This scalar match does not equate worst-case record protection, and uncharged preprocessing and proposals prevent an end-to-end privacy interpretation (*Fig. 1*).

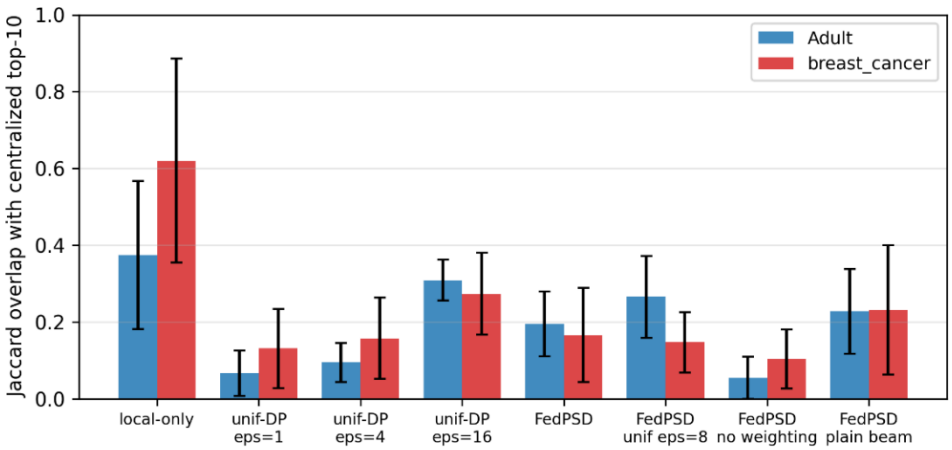

Fig. 2. Reported Jaccard overlap with the centralized depth-three heuristic search.

Local-only releases exact counts; uniform and personalized labels identify noise settings rather than verified end-to-end private pipelines (*Fig. 2*).

5.3 | Ablations: Where the Gain Comes From

Table 4 compares aggregation and proposal variants. Removing precision weighting lowers mean Adult WRAcc from 0.0503 to 0.0127 and breast-cancer WRAcc from 0.0559 to 0.0254. At the count-noise convention, budget-1 Adult clients have scale 222 while a budget of 6.25 gives scale about 35.5. Weighting attenuates noisy normalized rates but also changes population representation, so this ablation does not establish optimality or unbiasedness. Removing local proposals yields 0.0472 on Adult and 0.0598 on breast cancer; no paired significance or equivalence result is supplied for that ablation. Uniform budget 8 under the same machinery gives 0.0625 and 0.0721, at twice the scalar volume-weighted budget; whether that allocation is permissible depends on each client's constraints.

Table 4. Ablations over ten seeds (top-10 test WRAcc and Jaccard overlap with the centralized top-10).

Variant	Adult: test WRAcc	Adult: Jaccard	bc: test WRAcc	bc: Jaccard
FedPSD (full: personalised eps + weighting + privacy beam)	0.0503 ± 0.0177	0.196 ± 0.084	0.0559 ± 0.0294	0.166 ± 0.122
Uniform eps = 8, same machinery (double volume budget)	0.0625 ± 0.0149	0.266 ± 0.107	0.0721 ± 0.0217	0.148 ± 0.079
No quality weighting (plain-sum aggregation)	0.0127 ± 0.0142	0.055 ± 0.055	0.0254 ± 0.0375	0.104 ± 0.077
Plain federated beam (no local proposals)	0.0472 ± 0.0171	0.228 ± 0.110	0.0598 ± 0.0261	0.232 ± 0.168

5.4 | The Privacy-Utility Frontier

Table 5 reports quality under nominal allocation settings. Adult's personalized WRAcc of 0.0503 recovers about 29.6% of the gap between uniform 4 (0.0346) and the centralized reference (0.0877); the corresponding breast-cancer fraction is about 10.3% using rounded means. It reaches 71.0% of Adult uniform-16 quality and about 64.3% on breast cancer. Linear interpolation between uniform budgets 4 and 16 gives approximately 9.2 for Adult, but log-axis interpolation gives a different value and neither is a measured effective privacy budget. The two-corpus difference is consistent with size skew, but no factorial experiment separates size skew, population heterogeneity, budget dispersion, and estimand change.

Table 5. Reported allocation-quality comparison using nominal volume-weighted budgets.

Policy	Volume-Weighted eps	Adult: Test WRAcc	bc: Test WRAcc	Adult: % of Centralized	bc: % of Centralized
Uniform-DP	1	0.0166 ± 0.0169	-0.0042 ± 0.0506	19.0%	-3.6%
Uniform-DP	4	0.0346 ± 0.0080	0.0488 ± 0.0264	39.4%	41.4%
FedPSD (personalised)	4	0.0503 ± 0.0177	0.0559 ± 0.0294	57.3%	47.4%
Uniform eps = 8, FedPSD machinery	8	0.0625 ± 0.0149	0.0721 ± 0.0217	71.3%	61.2%
Uniform-DP	16	0.0708 ± 0.0063	0.0870 ± 0.0184	80.7%	73.9%

Note: Percentages express quality relative to the centralized heuristic reference; they do not equate worst-case privacy protection.

5.5 | Statistical Analysis

The source reports two-sided Wilcoxon tests on twenty seed-by-corpus blocks [25], Friedman over four method labels [26], and Holm correction over three contrasts [27]; Demsar's independent-dataset comparison framework [28] does not turn these resamples into twenty independent datasets. The pooled FedPSD-uniform4 contrast is W=50, p=0.040, with 16 of 20 favourable blocks, while the reported Friedman statistic is 50.85. These are exploratory summaries conditional on two fixed corpora and repeated, overlapping splits. The pooled p-value is not an Adult-specific result, and preregistration, raw pairing, ties/zeros and test implementation require documentation.

Table 6. Statistical analysis: paired Wilcoxon signed-rank tests on top-10 test WRAcc over 20 seed-by-dataset blocks (ten seeds x two corpora), with Holm correction over the three pairwise tests; Friedman chi-square(3) = 50.85, p = 5.3e-11.

Comparison (top-10 Test WRAcc, 20 Blocks)	W	p	Holm p	Outcome
FedPSD vs. centralized	0.0	1.91e-06	5.72e-06	Centralized better
FedPSD vs. local only	3.0	9.54e-06	1.91e-05	Local-only better
FedPSD vs. uniform 4	50.0	4.00e-02	4.00e-02	FedPSD better

5.6 | Findings, Honest Losses, and Scope

At the scalar volume-weighted budget convention, personalized allocation with precision weighting has higher mean WRAcc than uniform4 on both reported corpora, most clearly on Adult. The pooled W=50, p=0.040 does not establish a corpus-specific significance test or a causal mechanism for skew. At budget 1 Adult mean WRAcc remains positive at 0.0166, while breast-cancer mean is -0.0042; the claim that both corpora fall below baseline is unsupported. Statistics sent rise from 1170 to 1245 on Adult and 706 to 826 on breast cancer, or about 6.4% and 17.0%; these counts do not represent comparable full network bytes across raw-record and count-based baselines.

Local-only has higher mean WRAcc than every noisy federated variant and releases exact selected counts; it is a nonprivate reference with an undercount convention, not a valid privacy baseline. The centralized depth-three beam gives the highest observed mean here, but is not a mathematical upper bound. Uniform16 and the same-machinery uniform8 ablation show that allocation constraints matter. Adult FedPSD standard deviation is 2.21 times uniform4's (0.0177/0.0080), implying about 4.90 times the variance, not twice the variance. Routing and population correction are untested; private preprocessing, proposal selection, data-derived allocations, and transcript-wide composition remain unresolved.

The source reports a small campaign dated 2026-10-11. Execution and regeneration require the experiment artifacts; the present manuscript provides summary evidence rather than an independently reproduced result.

6 | Discussion: Limitations and Counterarguments

6.1 | Limitations

The measured variant omits privacy charges for description proposals and pooled preprocessing, and does not establish the declared whole-record budgets. Candidate caps, baseline release, fixed-size adjacency, zero-client handling, noisy-count projection, and public allocation inputs must be audited. Precision weighting changes the population represented by discovered subgroups. The centralized heuristic reference, unequal description depths, two corpora, and overlapping resamples also restrict interpretation. Client-level institutional protection remains outside the design.

6.2 | Alternative Viewpoints and Responses

A uniform budget can simplify governance and may be required by client-specific constraints; personalized budgets permit alternative allocations when those clients explicitly authorize them [8]. Formal guarantees depend on complete release accounting rather than an equitable-treatment slogan [9]. Syntactic anonymity and differential privacy address different threat models [29], [30]. Symbolic descriptions may assist governance, but the current campaign attaches noisy point scores rather than certified quality bounds.

If the complete mechanism satisfies the client-specific guarantees, a record at client i is protected at its own E_i, and a uniform worst-case statement uses $\max_i E_i$. Publishing a budget vector clarifies this heterogeneity but cannot repair uncharged releases. The size-weighted mean is a reporting convention rather than an adversary's privacy parameter. Governance validity and authorization of the allocation require a real consortium study.

7 | Conclusion and Future Directions

FedPSD studies heterogeneous count-noise allocation and precision-weighted subgroup discovery. In the reported campaign, mean held-out top-10 WRAcc rises from 0.0346 to 0.0503 on Adult and from 0.0488 to 0.0559 on breast cancer relative to uniform budget 4, with a pooled W=50,p=0.040 across twenty resampling blocks. The weighting ablation lowers Adult WRAcc to 0.0127, whereas uniform8 with the same machinery achieves 0.0625. These findings describe two corpora and changed mixture scoring. The complete pipeline is

not proven differentially private because data-dependent proposals, preprocessing, and allocations are uncharged; routing and post-stratification are untested.

The next stage must supply code and raw outputs, confirm outcome coding and empty-client treatment, audit actual count pairs against query caps, and privatize or fix preprocessing, proposal selection, baseline release, and allocation inputs. It should specify whether population or precision-mixture quality is the target, compare equal-depth searches, and evaluate routing and population correction independently. Tighter composition [7], [14] and cryptographic aggregation [21] can then be assessed under a complete transcript model.

Data Availability

UCI Adult and the scikit-learn breast-cancer benchmark are public [22], [23]. The reported federated partitions, budgets, splits, and derived statistics require the seeded procedure and raw outputs, including exp_23_results.json.

Code Availability

Reproduction requires exp_23.py, exp_23_results.json, the seed list and original figure-generation workflow. A stable repository or attached supplement is not identified in the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

Soheil Fakheri: problem formulation, methodology, and writing—original draft. Jyotsana Grover: software, validation, and review—editing. All authors read and approved the final manuscript.

Generative AI Disclosure

AI-assisted drafting tools were used in preparing the manuscript text under the direction and scientific responsibility of the authors; all methodological choices, claims, experimental results, and final content were determined and verified by the authors.

References

[1] Lavrač, N., Kavšek, B., Flach, P., & Todorovski, L. (2004). Subgroup discovery with CN2-SD. *Journal of machine learning research*, *5*(2), 153–188. https://dl.acm.org/doi/10.5555/1005332.1005338

[2] Atzmueller, M. (2015). Subgroup discovery. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, *5*(1), 35–49. https://doi.org/10.1002/widm.1144

[3] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-efficient learning of deep networks from decentralized data. *The 20th international conference on artificial intelligence and statistics* (pp. 1273–1282). PMLR. https://proceedings.mlr.press/v54/mcmahan17a.html

[4] Kairouz, P., & McMahan, H. B. (2021). Advances and open problems in federated learning. In *Foundations and trends® in machine learning* (pp. 1–210). Now Publishers Inc. https://doi.org/10.1561/2200000083

[5] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). *Federated learning with non-iid data*. https://doi.org/10.48550/arXiv.1806.00582

[6] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and trends®in theoretical computer science*, *9*(3–4), 211–487. https://doi.org/10.1561/0400000042

[7] Feldman, V., & Zrnic, T. (2021). Individual privacy accounting via a renyi filter. *Advances in neural information processing systems*, *34*, 28080–28091. https://proceedings.neurips.cc/paper_files/paper/2021/hash/ec7f346604f518906d35ef0492709f78-Abstract.html

[8] Jorgensen, Z., Yu, T., & Cormode, G. (2015). Conservative or liberal? personalized differential privacy. *2015 IEEE 31st international conference on data engineering* (pp. 1023–1034). IEEE. https://doi.org/10.1109/ICDE.2015.7113353

[9] Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *Proceedings of the 3rd innovations in theoretical computer science conference* (pp. 214-226). Association for Computing Machinery. https://doi.org/10.1145/2090236.2090255

[10] Novak, P. K., Lavrač, N., & Webb, G. I. (2009). Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of machine learning research*, *10*(2), 377–403. https://www.jmlr.org/papers/volume10/kralj-novak09a/kralj-novak09a.pdf

[11] Van Leeuwen, M., & Knobbe, A. (2012). Diverse subgroup set discovery. *Data mining and knowledge discovery*, *25*(2), 208–242. https://doi.org/10.1007/s10618-012-0273-y

[12] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* (pp. 265-284). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11681878_14

[13] Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., & Smith, A. (2011). What can we learn privately? *SIAM journal on computing*, *40*(3), 793–826. https://doi.org/10.1137/090756090

[14] Kairouz, P., Oh, S., & Viswanath, P. (2015). The composition theorem for differential privacy. *International conference on machine learning* (pp. 1376-1385). PMLR. https://proceedings.mlr.press/v37/kairouz15.html

[15] Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., … Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. *IEEE transactions on information forensics and security*, *15*, 3454–3469. https://doi.org/10.1109/TIFS.2020.2988575

[16] McMahan, H. B., Ramage, D., Talwar, K., & Zhang, L. (2017). *Learning differentially private recurrent language models*. https://doi.org/10.48550/arXiv.1710.06963

[17] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. *2017 IEEE symposium on security and privacy (sp)* (pp. 3–18). IEEE. https://doi.org/10.1109/SP.2017.41

[18] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*, 50–60. https://doi.org/10.1109/MSP.2020.2975749

[19] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. *Proceedings of machine learning and systems*, *2*, 429–450. https://proceedings.mlsys.org/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html

[20] Wang, H., Yurochkin, M., Sun, Y., Papailiopoulos, D., & Khazaeni, Y. (2020). *Federated learning with matched averaging*. https://doi.org/10.48550/arXiv.2002.06440

[21] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., … & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. *Proceedings of the 2017 acm sigsac conference on computer and communications security* (pp. 1175–1191). Association for Computing Machinery. https://doi.org/10.1145/3133956.3133982

[22] Kohavi, B., & Becker, R. (1996). *Adult [Dataset]*. https://doi.org/10.24432/C5XW20

[23] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., & Grisel, O. (2011). Scikit-learn: Machine learning in Python. *The journal of machine learning research*, *12*, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

[24] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., & Cournapeau, D. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

[25] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, *1*(6), 80–83. https://doi.org/10.2307/3001968

[26] Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, *32*, 675–701. https://doi.org/10.1080/01621459.1937.10503522

[27] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, *6*(2), 65–70. https://www.jstor.org/stable/4615733

[28] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of machine learning research*, *7*(1), 1–30. http://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf

[29] Aggarwal, C. C., & Philip, S. Y. (2008). *Privacy-preserving data mining: models and algorithms*. Springer Science & Business Media. https://www.aiu.edu.sy/public/en/electronic-book/privacy-preserving-data-mining-models-and-algorithms

[30] Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, *10*(05), 557–570. https://doi.org/10.1142/S0218488502001648