Multi-Agent *Experiment Design*

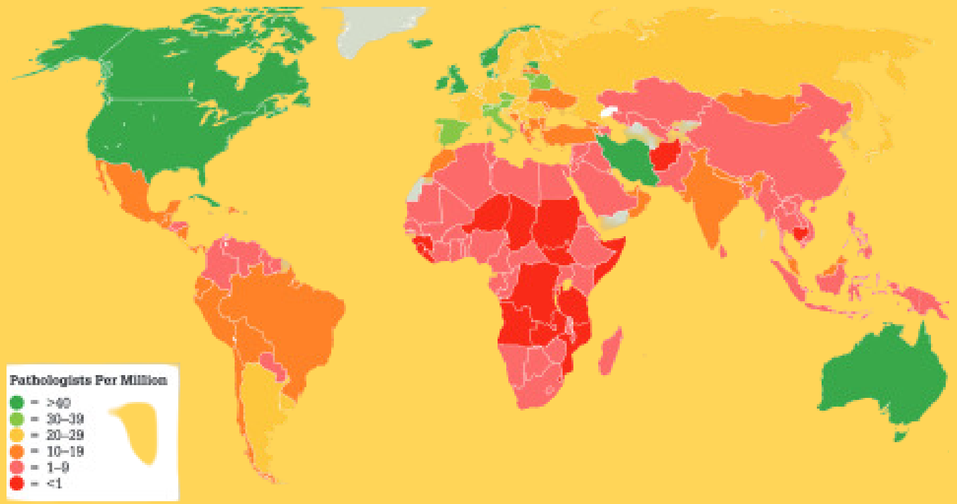




"Evaluating and Incentivizing Diverse Data Contributions in Collaborative Learning"

Baihe Huang, **Sai Praneeth Karimireddy**, Michael Jordan

Shortage of Cancer expertise



> **"Cancer patients face life-threatening delays due to lack of staff, say UK radiologists."**
>
> — The Guardian, 8 Jun 2023

- Global shortage of expertise to diagnose and manage cancer even in high-income regions.
- Life-threatening consequences.
- AI can maybe help?

Cancer Registry of Norway

Data scarcity & diversity

- Very little data. 562 cases of Sarcoma in 2022 across Norway.
- Worse for marginal groups. No dark-skinned skin cancer data.
- True in the US as well. Gets dire in low-income regions.





"There are two big areas... One is data systems..The second is ...reaching populations that have been left out."
- Bertagnolli 2024 (new NIH Chief)

Cancer Registry of Norway

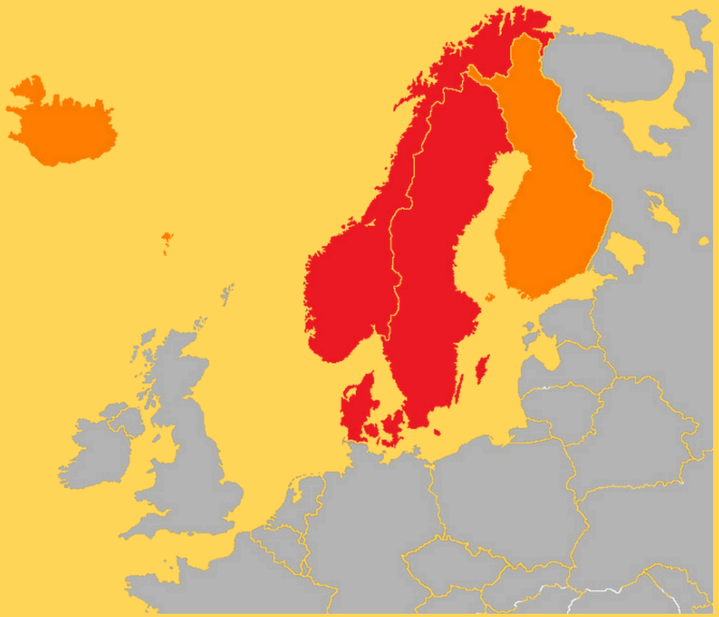
NORDCAN Collaboration

Passive users to active agents

7+ countries and 21 registries want to collaborate.

But,

- Strict data regulations. Use collaborative learning.
- **Beneficence:** risk-benefit analysis for each sub-population. Need econ/game theory.



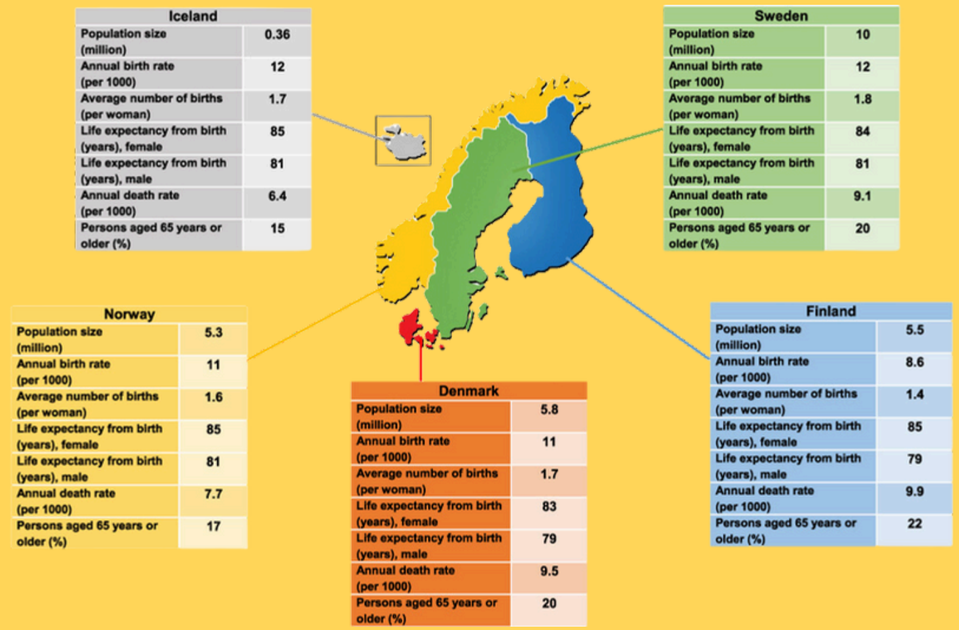
Cancer Registry of Norway

Position paper with Norway cancer registry

[*Karimireddy* et al. FMEC 2023]

Principal-Agents Formalism

- ***n*** countries (agents), each can ***freely choose***
  - ***how much*** data
  - ***which*** samples
- Data incurs real costs (privacy, side effects)
- Efficiency critical, but over which subpopulation?
- This is a multi-agent problem.



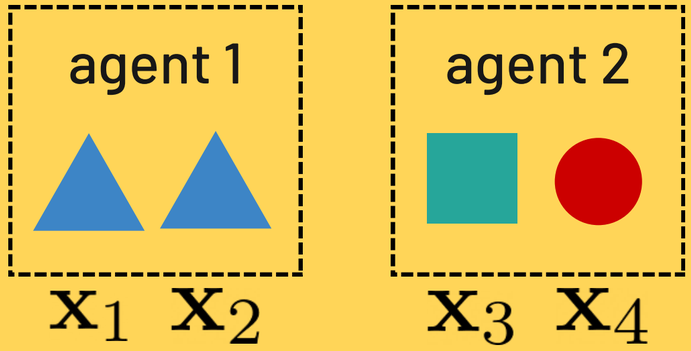
[*Karimireddy*, Guo, Jordan. **best paper** FL-NeurIPS 2022]

[Huang et al. 24. Evaluating and Incentivizing Diverse Data Contributions in Collaborative Learning]

Principal-Agent Experiment Design

Heterogeneous populations

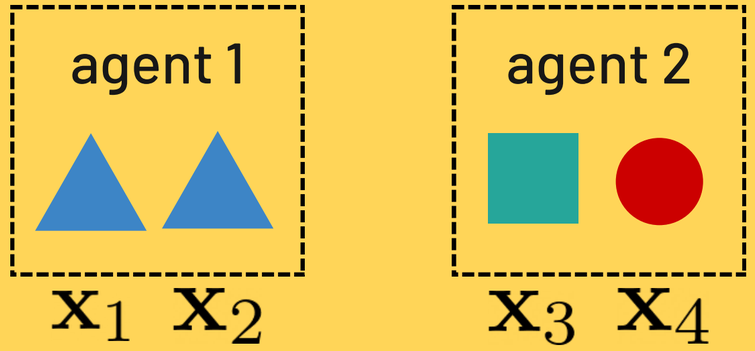
- Data types
  - **type** encoded by $\mathbf{x}$ = (30 yrs, Male, Asian, ...)
  - Total population has many types:
    $$\mathcal{X} = (\mathbf{x}_j)_{j=1,\ldots}$$
  - Partitioned between ***n*** agents. Agent i only has subtypes $\mathcal{X}_i \subseteq \mathcal{X}$.



Principal-Agent Experiment Design

Sampling strategy

- Data types
  - Total population has many types:
    $$\mathcal{X} = (\mathbf{x}_j)_{j=1,\dots}$$
  - Partitioned between ***n*** agents. Agent i only has subtypes $\mathcal{X}_i \subseteq \mathcal{X}$.
- Find the “best” sampling strategy:
  - m_j is **number of samples** of type $\mathbf{x}_j$.
  - Agent i shares $\mathbf{m}_i = (m_j, \text{ for } \mathbf{x}_j \in \mathcal{X}_i)$



$$\mathbf{m} = (2.3,\ 0.7,\ 1.8,\ 0.1)$$
$$= (\mathbf{m}_1,\ \mathbf{m}_2)$$

Classic Experiment Design

Estimating error

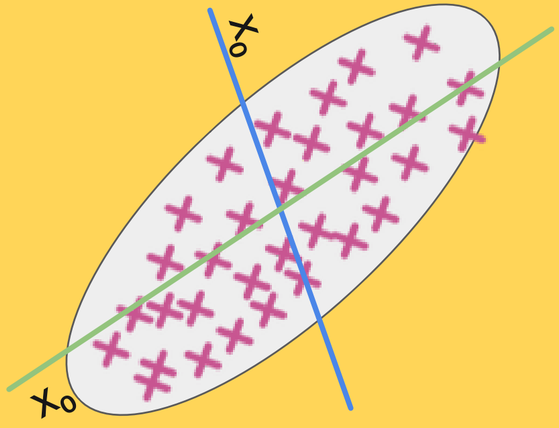
- Data space $\mathcal{X}$
- Total data sampled $\mathbf{m}$
- Agent i has types $\mathcal{X}_\mathbf{i}$
- Contributes $\mathbf{m}_i = (m_j \,|\mathbf{x}_j \in \mathcal{X}_i)$

- Error on any test $\mathbf{x_0}$ determined by $\mathcal{I}(\mathbf{m})$

$$\mathcal{E}(\mathbf{x_0}, \mathbf{m}) = \mathbf{x_0}^\top (\sum_{j=1}^{k} m_j \mathbf{x_j} \mathbf{x_j}^\top)^\dagger \mathbf{x_0}$$
$$= \mathbf{x_0}^\top \mathcal{I}(\mathbf{m})^\dagger \mathbf{x_0}$$

- Information matrix $\mathcal{I}(\mathbf{m})$ captures error on all possible test points.

Error determined by information matrix



[Wald 1943; Pukelsheim 2006]

Classic Experiment Design

Choosing a metric

- Data space $\mathcal{X}$
- Total data sampled $\mathbf{m}$
- Agent i has types $\mathcal{X}_\mathbf{i}$
- Contributes $\mathbf{m}_i = (m_j \,|\mathbf{x}_j \in \mathcal{X}_i)$

- Error on any test $\mathbf{x_0}$ determined by $\mathcal{I}(\mathbf{m})$

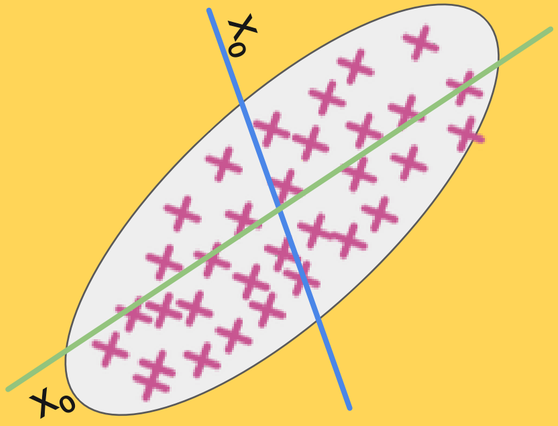
$$\mathcal{E}(\mathbf{x_0}, \mathbf{m}) = \mathbf{x_0}^\top (\sum_{j=1}^{k} m_j \mathbf{x_j} \mathbf{x_j}^\top)^\dagger \mathbf{x_0}$$
$$= \mathbf{x_0}^\top \mathcal{I}(\mathbf{m})^\dagger \mathbf{x_0}$$

- Metric compresses into a scalar error

$$\mathcal{E}_f(\mathcal{X}, \mathbf{m}) = \mathbf{f}\left(\mathcal{I}(\mathbf{m})^\dagger\right)$$

- $\mathbf{f}$ can be
  - Tr (A criterion)
  - log det (D criterion)
  - λmin (E criterion)
  - ...

Error determined by information matrix



[Wald 1943; Pukelsheim 2006]

Principal-Agent Experiment Design

Efficiency games

- Data space $\mathcal{X}$:
- Total data sampled $\mathbf{m}$
- Agent i has types $\mathcal{X}_\mathbf{i}$
- Contributes $\mathbf{m}_i = (m_j \,|\mathbf{x}_j \in \mathcal{X}_i)$

Global: $$\min_{\mathbf{m}} \underbrace{\mathcal{E}_f(\mathcal{X}, \mathbf{m})}_{\text{net error}} + \underbrace{c\|\mathbf{m}\|_1}_{\text{net cost}}$$

Agents: $$\min_{\mathbf{m}_i} \underbrace{\mathcal{E}_f(\mathcal{X}_i, (\mathbf{m}_i, \mathbf{m}_{\neg i}))}_{\text{agent error}} + \underbrace{c\|\mathbf{m}_i\|_1}_{\text{agent cost}}$$

Global and local objectives misaligned => inefficiency => ***Price of Anarchy***



Principal-Agent Experiment Design

Estimating agent's error

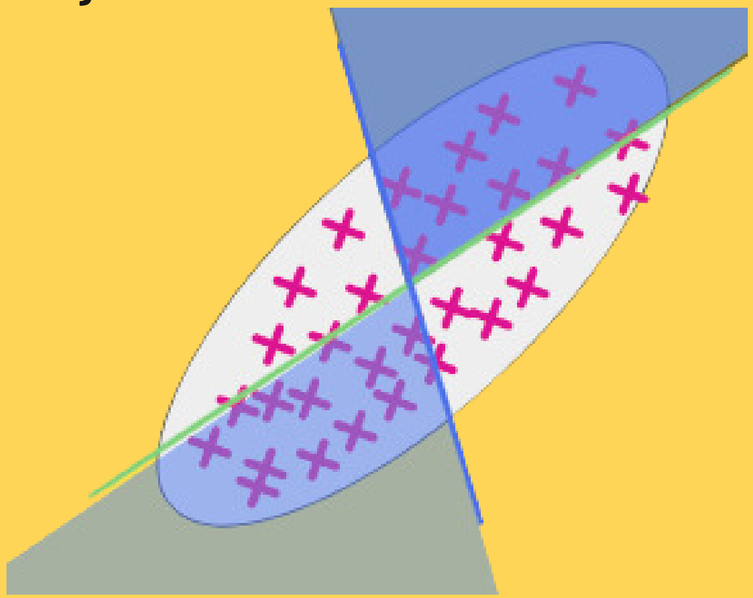
- Data space $\mathcal{X}$
- Total data sampled $\mathbf{m}$

- Agent i has types $\mathcal{X}_\mathbf{i}$
- Contributes $\mathbf{m}_i = (m_j \,|\mathbf{x}_j \in \mathcal{X}_i)$

- Agent i only cares about errors on $\mathcal{X}_\mathbf{i}$

$$\mathcal{E}_f(\mathcal{X}_i, (\mathbf{m}_i, \mathbf{m}_{\neg i})) = \mathbf{f}\left(\Pi_i \mathcal{I}(\mathbf{m})^\dagger \Pi_i\right)$$

Projected information matrix



Principal-Agents Formalism

Efficient data sampling strategies

Is the sampling strategy incentive-compatible?

- *Not minimizing avg error:* Global efficiency prioritizes majority agents 1, 2 who have same type vs. minority agent 3 who has diff type.



Principal-Agents Formalism

Efficient data sampling strategies

Is the sampling strategy incentive-compatible?

- *Not minimizing avg error*
- *Not minimizing worst-case error:*
  - Agent 2 has a harder task.
  - Global efficiency means 95% acc on all.
  - But agent 2 has 2x data costs.
  - Likely prefers lower acc for lower data costs.

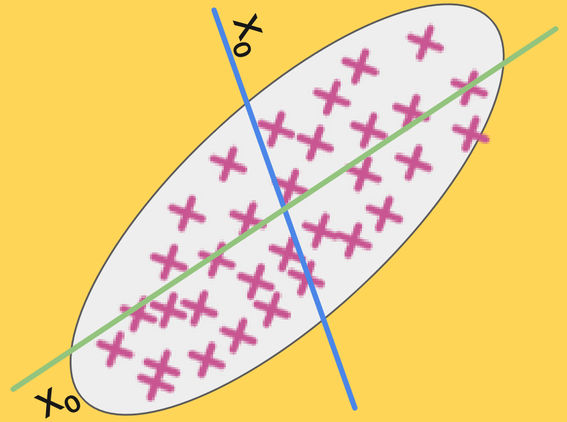


Principal-Agent Experiment Design

Our Theorem

- For **f = log det** (D-criterion),
  - unique equilibrium
  - always Price of Anarchy = 1!
  - invariant to population splits.
- Not true for others: avg (A, V), nor min-max (G, E).

$$\log\det(\Pi\,\mathcal{I}(m)) = \log\det(\mathcal{I}(m)) + c$$



Upshot: Incentive-compatible only if everyone uses D-criterion. So, use it.

[Huang et al. 24. Evaluating and Incentivizing Diverse Data Contributions in Collaborative Learning]

Principal-Agent Experiment Design

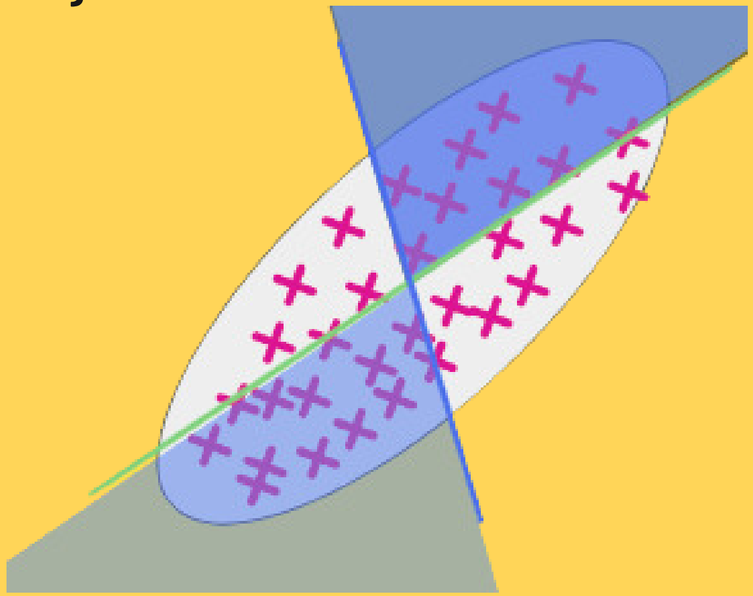
Free-riding games

- Data space $\mathcal{X} = (\mathbf{x_1}, \ldots, \mathbf{x_k})$
- Total data $\mathbf{m} = (m_1, \ldots, m_k)$
- Agent i has types $\mathcal{X}_\mathbf{i}$
- Contributes $\mathbf{m}_i = (m_j \,|\mathbf{x}_j \in \mathcal{X}_i)$

Agents: $\min_{\mathbf{m}_i} \mathcal{E}_f(\mathcal{X}_i, (\mathbf{m}_i, \mathbf{m}_{\neg i})) + c\|\mathbf{m}_i\|_1$

- Agents may end up with very different utilities
- Others' data is free => unequal net costs => **free-riding**.
- Add penalties for not pulling your weight.

Projected information matrix



Principal-Agent Experiment Design

Our Theorems

1. Fairness is incompatible with efficiency.

2. A simple penalization scheme guarantees
   a) optimal data contributions, and
   b) equal net utility for all.

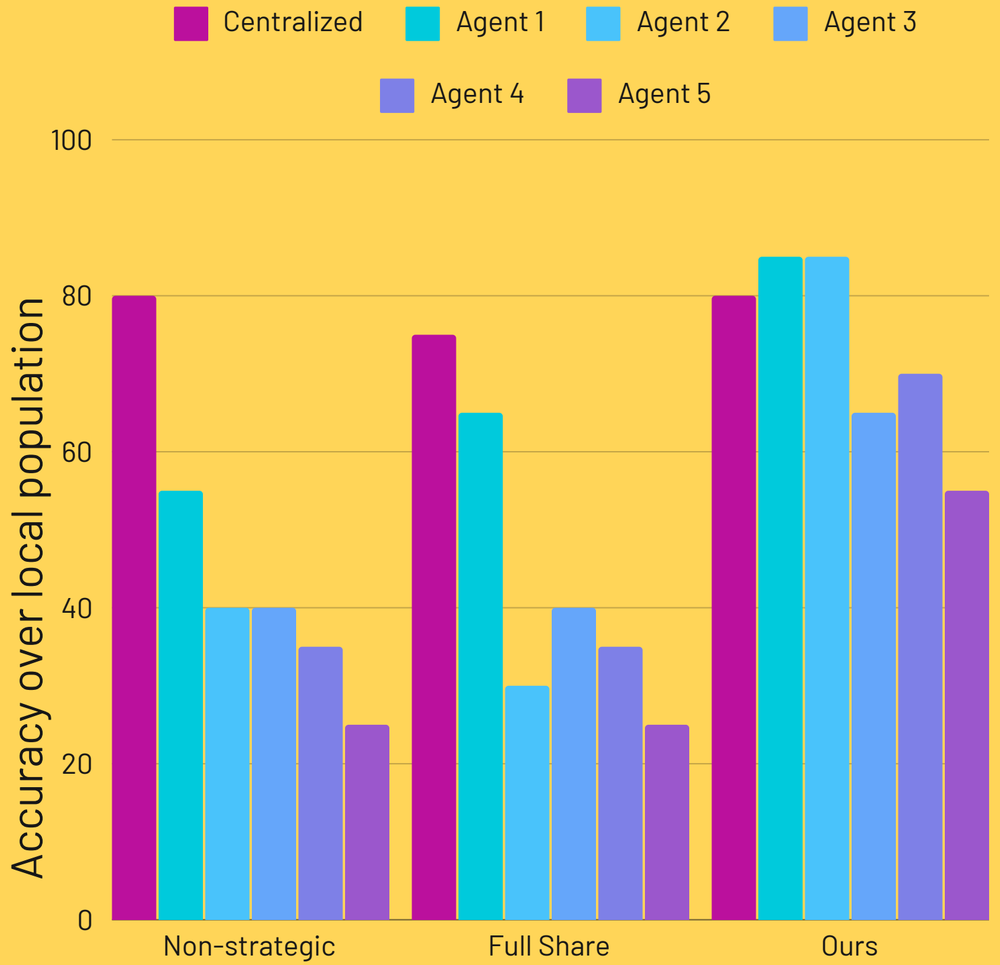
3. Can trade-off efficiency/fairness.



Principal-Agent Experiment Design

Our Results

- Finetuning ResNet-18 on MedMNIST
- Types = classes. Split heterogeneously using Dirichlet sampling.
- All work if centralized, o.w.
  - Non-strategic: inefficient
  - Full sharing: unequal
  - Ours: efficient and “fair”



Conclusion

- Main takeaways:
  - log-det criterion is the only efficient criterion in multi-agent exp design.
  - simple penalization scheme can trade off fairness and efficiency.
- Call to action: Multi-agent versions of all learning/optimization/stats problems.

[Huang et al. 24. Evaluating and Incentivizing Diverse Data Contributions in Collaborative Learning]