Results - Simulation Data

- Dummy example: study the association between medication status and fetal loss

Covariate	How is data generated?
medication status (yes/no)	Bernoulli(p)
race (white; non-white)	Bernoulli(p)
weight, age	Normal(μ,σ)

- Outcome generated by:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4, \quad \boldsymbol{\beta} \sim U(-1, 1)$$

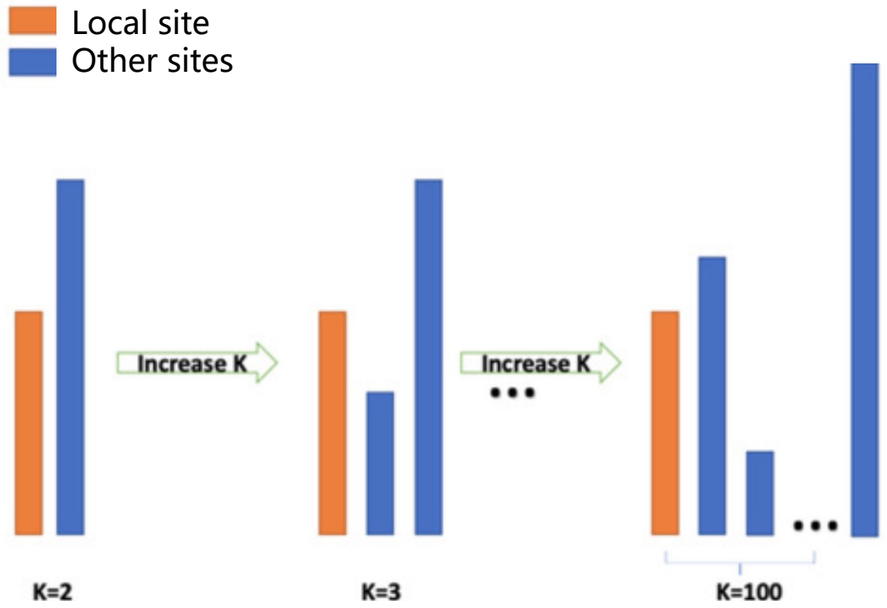
Results - Simulation Settings

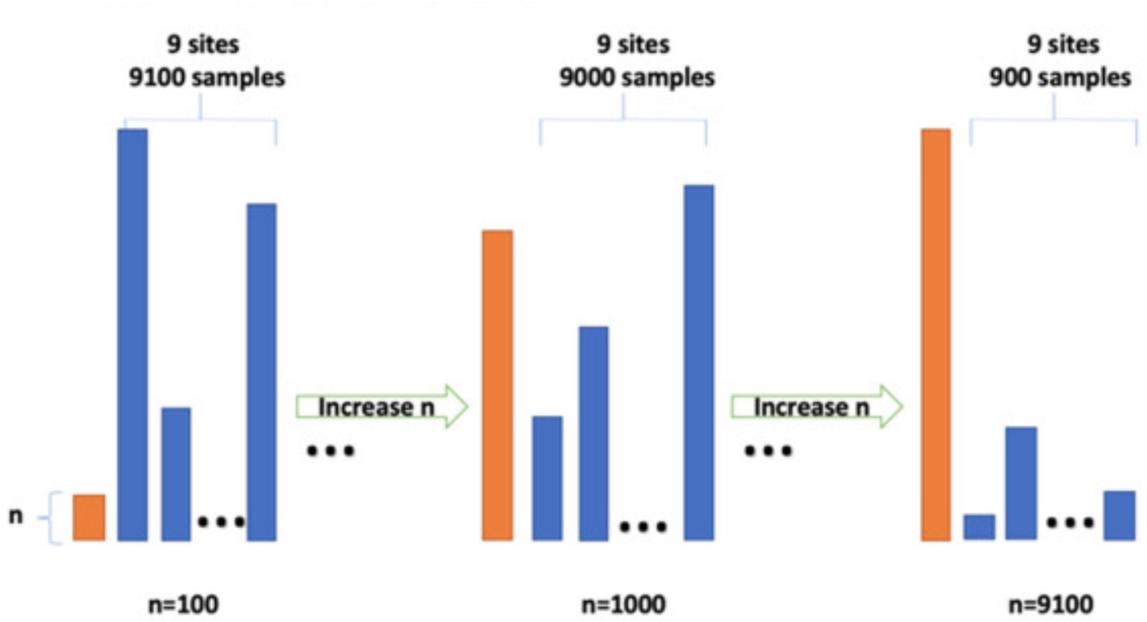
Setting A:

- Local sample size fixed
- Number of total sites grows

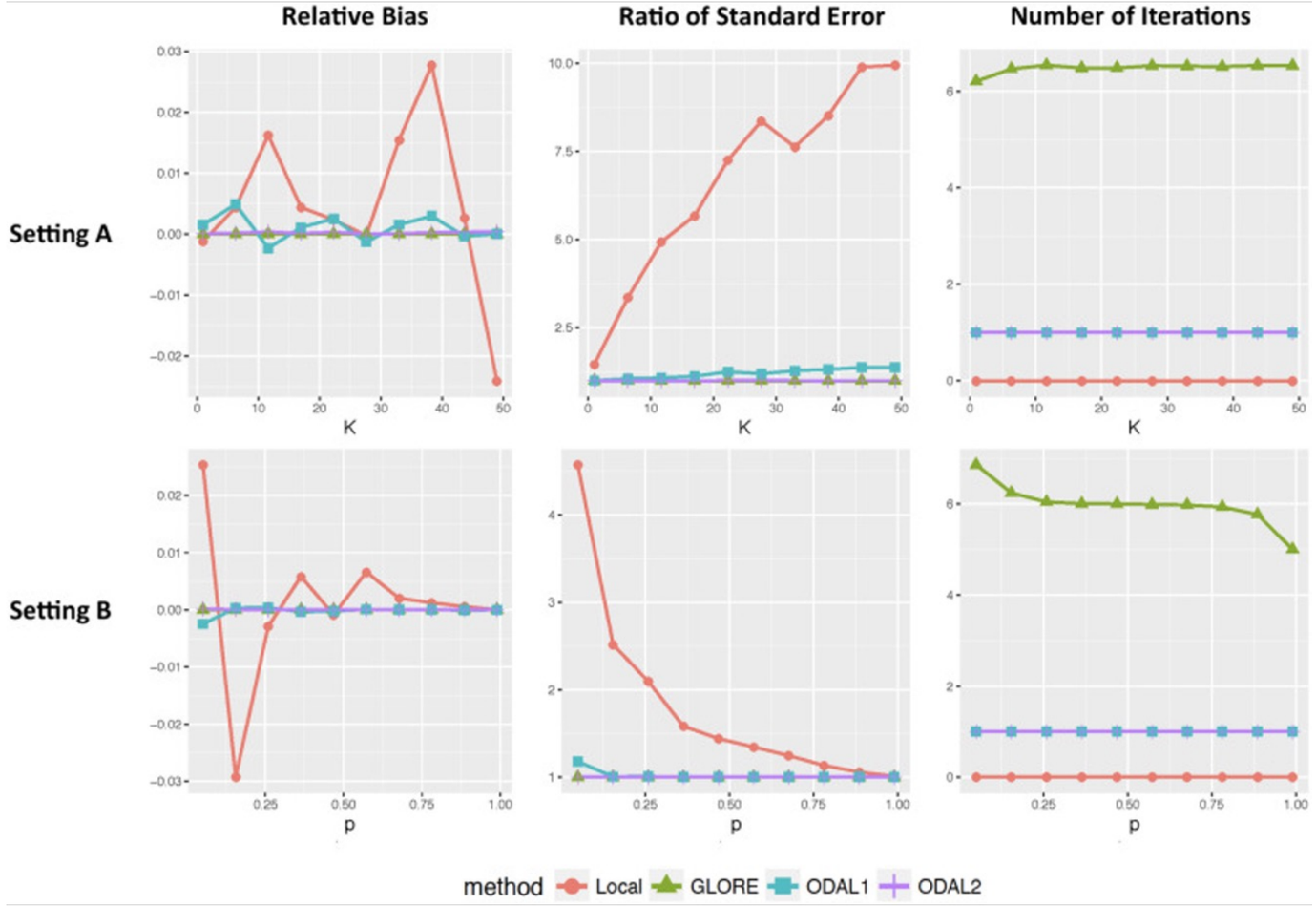
Setting B:

- Local sample size grows
- Number of total sites fixed





Results - Simulation Output



Key points

- **Accuracy**

  GLORE > ODAL2 > ODAL1

- **Communication cost**

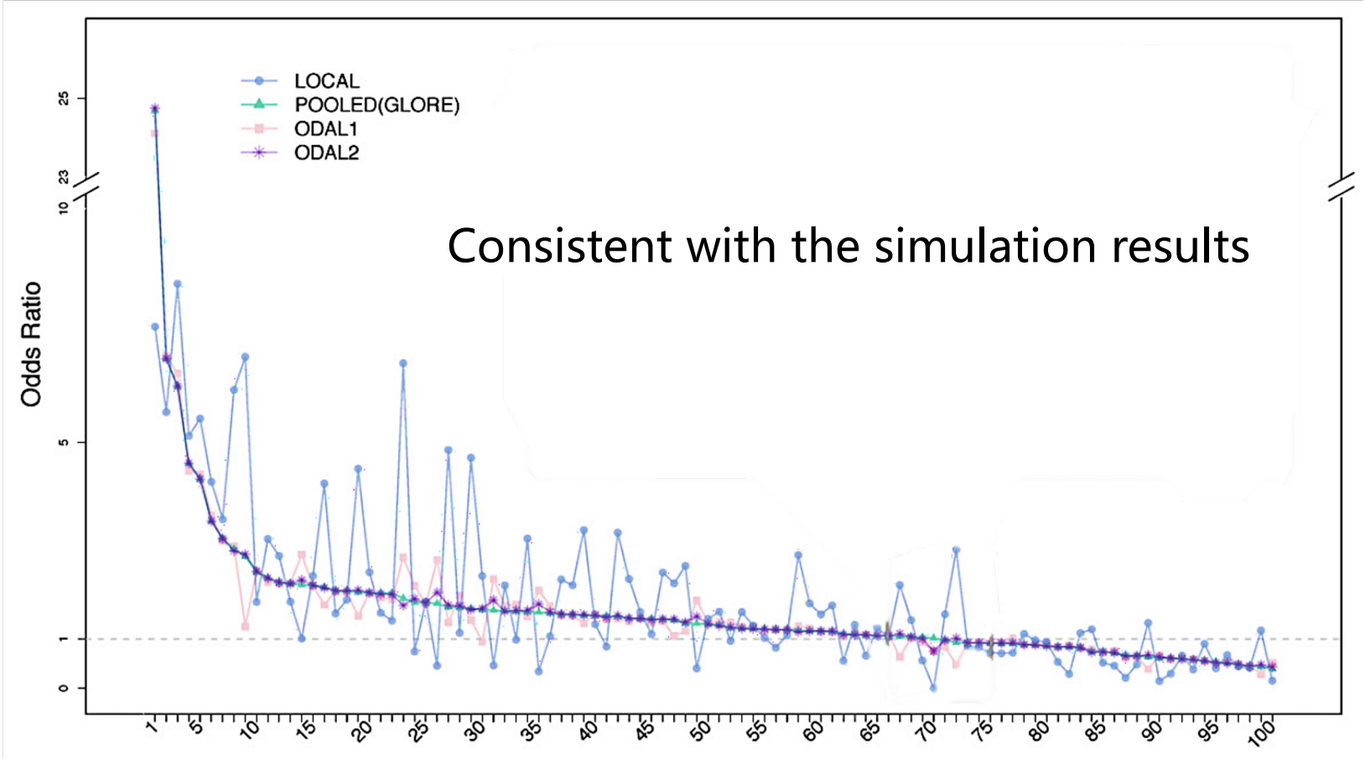
  GLORE > ODAL2> ODAL1

Results - Real Clinical Study

Data

- Samples: ~35k
- Ten sites: 10% to mimic local site
- Exposure(s): 100 medications
- Covariates: age, weight, race, BMI

Results



Conclusion

In summary, ODAL2:

- Provides accurate estimates;
- Communication-efficient

In practice, tradeoff between accuracy and communication cost

Assess data complexity before choosing an algorithm!!

- Number of sites

- Dimension of feature space