Penn Medicine

Department of Biostatistics, Epidemiology and Informatics

Padé approximant meets federated learning: a nearly lossless, one-shot algorithm for evidence synthesis

Qiong Wu, Postdoctoral Researcher, University of Pennsylvania

Joint work with Yong Chen, Martijn Schuemie, Marc Suchard, Patrick Ryan, George Hripcsak, and Charles Rohde

Presented at the OHDSI Community Call

September 26th, 2023

PennCIL
A Computing · Inference · Learning lab at University of Pennsylvania
https://penncil.med.upenn.edu/

Distributed Research Networks (DRNs)

pcornet® The National Patient-Centered Clinical Research Network

Nemours.

The Children's Hospital of Philadelphia®

Nationwide Children's

Boston Children's Hospital

PEDSNet CDRN

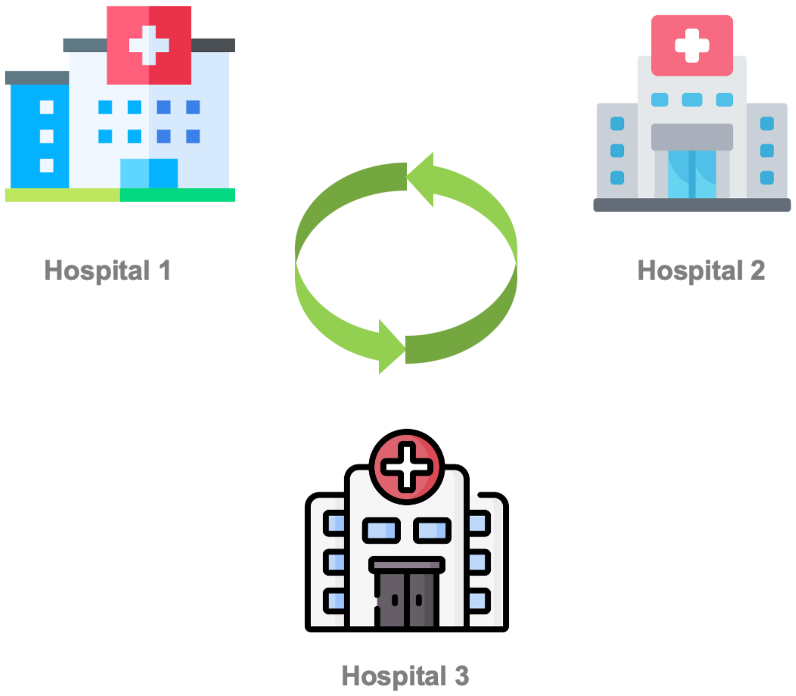
Cincinnati Children's®

Children's Hospital • St. Louis

Children's Hospital Colorado®

Seattle Children's

Generate evidence from DRNs



- Benefits:
  - Generalizable findings
  - Larger amount of data (better statistical power)
  - **Easier to study rare events:**
    - Adverse event from drugs: important in pharmacovigilance and pharmacoepidemiology
    - Rare disease: vasculitis in PCORNet

- Challenges:
  - Protection of patients' privacy
  - Communication-efficient
  - Unique challenge in studying rare diseases

picture from internet

Example: rare adverse effects

- Goal: comparing depression drugs on rare adverse effects using observational healthcare databases.
- Four databases
  - IBM MarketScan Commercial Claims and Encounters (CCAE)
  - IBM MarketScan Medicare Supplemental Beneficiaries (MDCR)
  - IBM MarketScan Multi-state Medicaid (MDCD)
  - Optum's de-identified Clinformatics Data Mart database
- Comparisons
  - Comparison 1: amitriptyline (target treatment) vs. citalopram (comparator treatment) as risk factors for the occurrence of acute liver injury. There are non-zero counts across all four databases in this comparison.
  - Comparison 2: nortriptyline (target treatment) and duloxetine (comparator treatment) for the risk of acute liver injury, where two databases had zero counts.
  - Comparison 3: nortriptyline (target treatment) and venlafaxine (comparator treatment) in terms of the risk of decreased libido in which two databases had zero counts in the target and highest counts in the comparator cohort.

Example: rare adverse effects

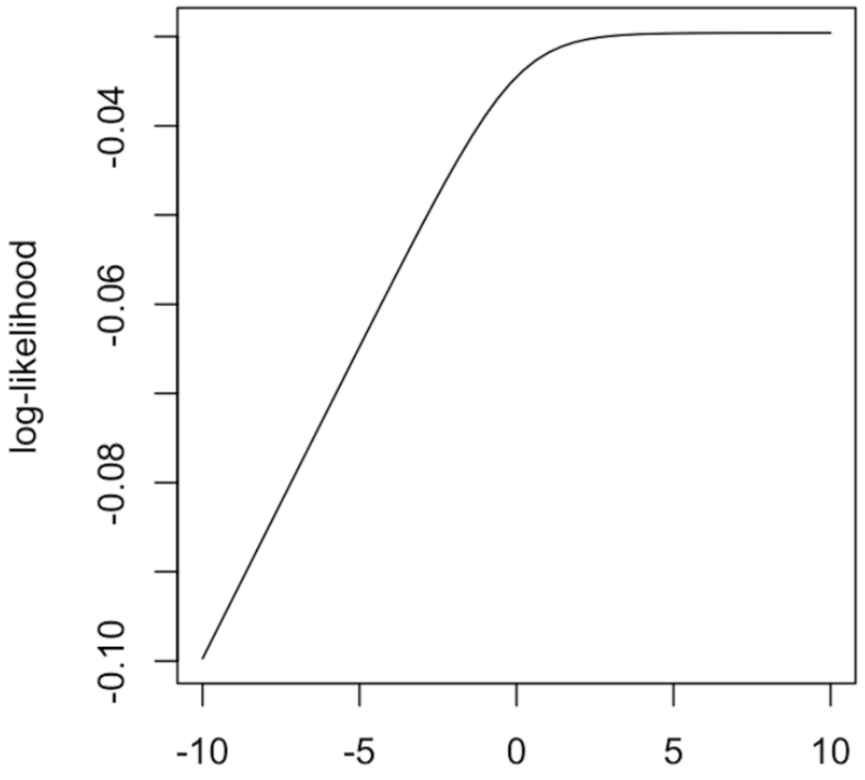
- A common approach: in each database *j*
  - Apply propensity score stratification to adjust for confounding variables including demographics, prior conditions, exposures, procedures, measurements, etc.
  - Stratified Cox proportional hazard model on each site *j*

$$\lambda(t|x_{ij}) = \lambda_{sj}(t)\exp(x_{ij}\theta)$$

where $\lambda_{sj}(t)$ is the baseline hazard function of the *s*-th stratum in *j*-th site.

- Goal: estimate treatment effect θ collaboratively using multiple databases

Ill behaved likelihood for rare events



- With few or zero events, the likelihood of a database is **ill-behaved**.
  - Monotone likelihood of Cox regression (Heinze et al., 2001, Nagashima et al., 2017)
- Meta-analysis can have a substantial bias in this case

Key idea

Communicate the local likelihood of each database and combine them at a master site

Heinze, G., & Schemper, M. (2001). A solution to the problem of monotone likelihood in Cox regression. *Biometrics*, *57*(1), 114-119.
Nagashima, K., & Sato, Y. (2017). Information criteria for Firth's penalized partial likelihood approach in Cox regression models. Statistics in medicine.

Communicate local likelihoods

- Idea 1: communicate local likelihoods on grids (of θ)? Costly!
- Idea 2: approximate local likelihoods with simplified functions and communicate the parameters?

Approximation theory

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Original Research Article

Combining cox regressions across a heterogeneous distributed research network facing small and zero counts

Statistical Methods in Medical Research
2022, Vol. 31(3) 438–450

DOI: 10.1177/09622802211060518
journals.sagepub.com/home/smm
SAGE

Martijn J. Schuemie[1,2,3], Yong Chen[4], David Madigan[1,5], and Marc A. Suchard[1,3,6]

Quadratic approximation

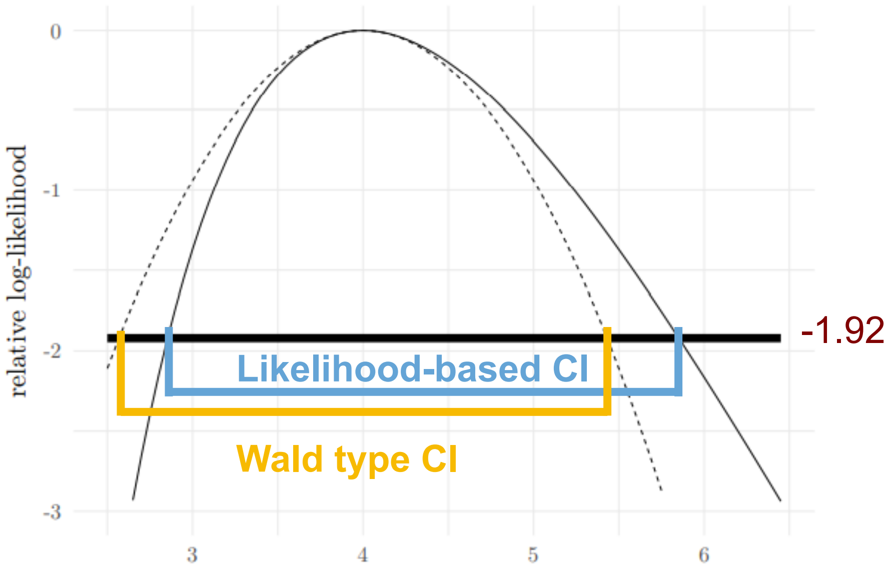
- Example: second-order Taylor expansion on MLE

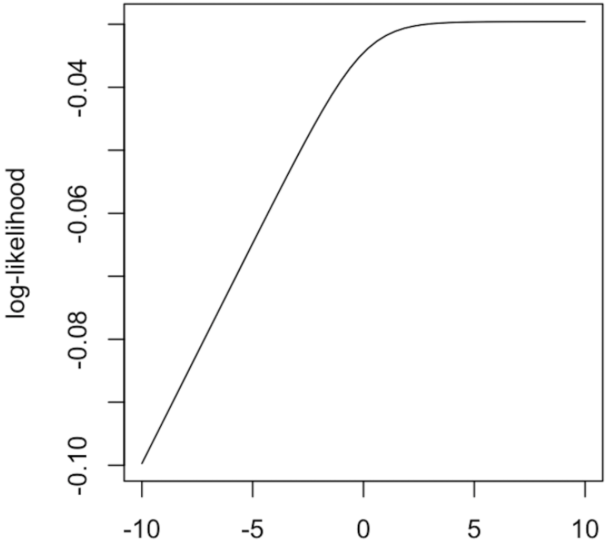
$$L_{Taylor}(\theta) \approx L(\hat{\theta}) + \nabla L(\hat{\theta})^T(\theta - \hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^T \nabla^2 L(\hat{\theta})(\theta - \hat{\theta})$$

- With quadratic approximation on the likelihood function, the likelihood-based confidence interval

$$2(L_{Taylor}(\theta) - L_{Taylor}(\hat{\theta})) \sim \chi^2$$

results in a Wald-type CI.





Padé approximants

WIKIPEDIA
The Free Encyclopedia

Create account Log in

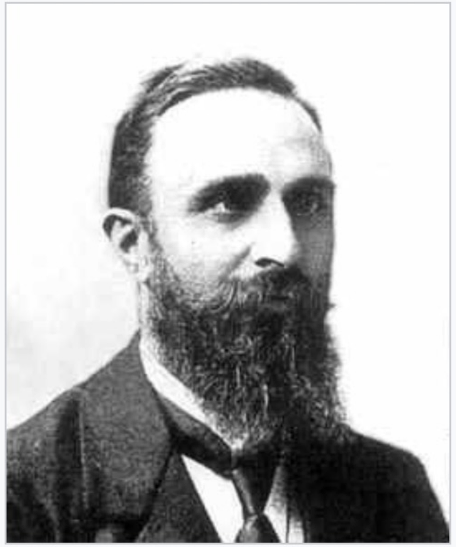
Padé approximant

17 languages

Article Talk Read Edit View history

From Wikipedia, the free encyclopedia

In mathematics, a **Padé approximant** is the "best" approximation of a function near a specific point by a rational function of given order. Under this technique, the approximant's power series agrees with the power series of the function it is approximating. The technique was developed around 1890 by Henri Padé, but goes back to Georg Frobenius, who introduced the idea and investigated the features of rational approximations of power series.



Henri Padé

The Padé approximant often gives better approximation of the function than truncating its Taylor series, and it may still work where the Taylor series does not converge. For these reasons Padé approximants are used extensively in computer calculations. They have also been used as auxiliary functions in Diophantine approximation and transcendental number theory, though for sharp results ad hoc methods— in some sense inspired by the Padé theory— typically replace them. Since Padé approximant is a rational function, an artificial singular point may occur as an approximation, but this can be avoided by Borel–Padé analysis.

Padé approximants

- Univariate Padé approximant

$$L_{\text{Padé}}(\beta) = \frac{a_0 + a_1(\beta - \bar{\beta}) + a_2(\beta - \bar{\beta})^2 + \cdots + a_m(\beta - \bar{\beta})^m}{1 + b_1(\beta - \bar{\beta}) + b_2(\beta - \bar{\beta})^2 + \cdots + b_n(\beta - \bar{\beta})^n}$$

$$L_{\text{Padé}}(\bar{\beta}) = L(\bar{\beta})$$

$$L^{(1)}_{\text{Padé}}(\bar{\beta}) = L^{(1)}(\bar{\beta})$$

$$L^{(2)}_{\text{Padé}}(\bar{\beta}) = L^{(2)}(\bar{\beta})$$

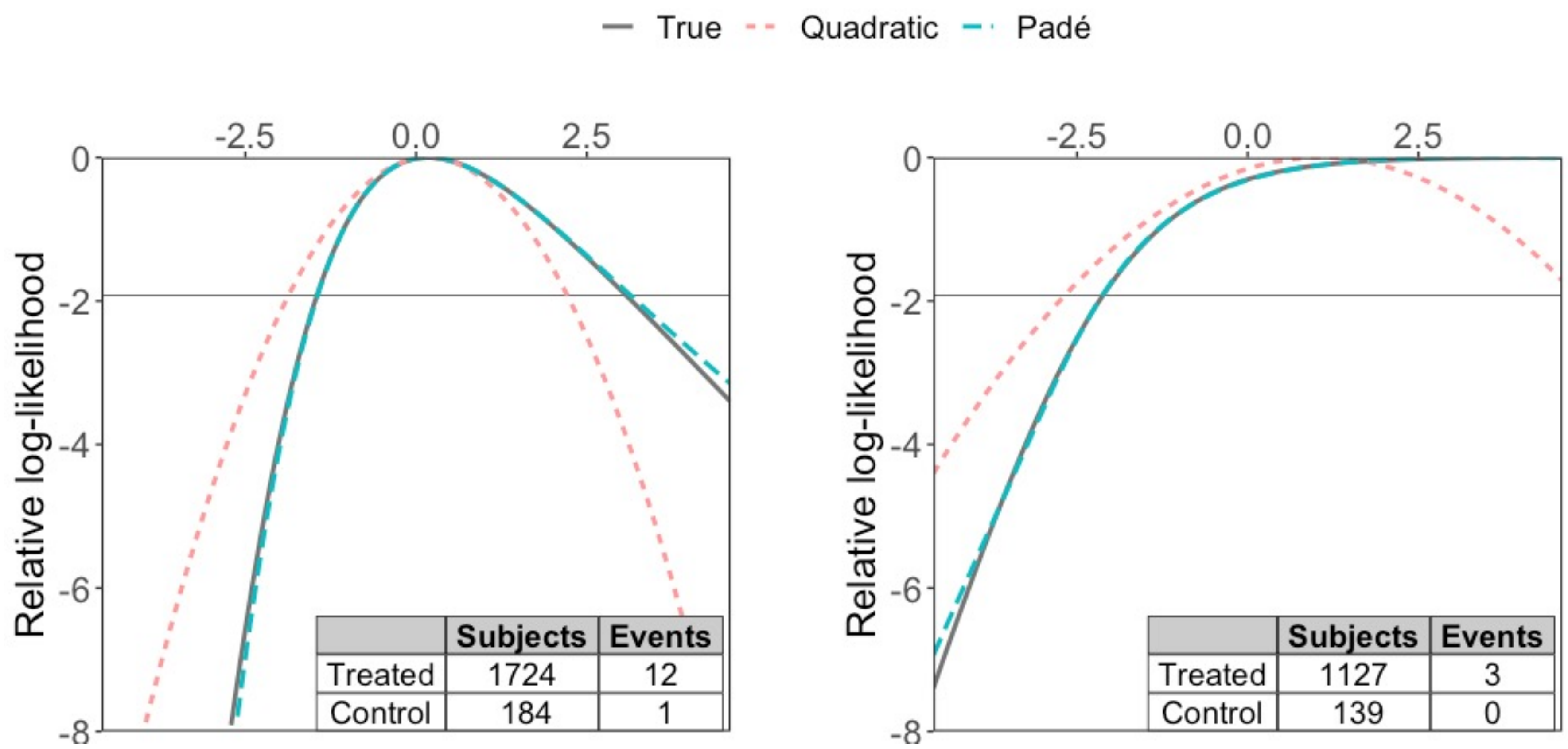
$$\cdots$$

$$L^{(m+n)}_{\text{Padé}}(\bar{\beta}) = L^{(m+n)}(\bar{\beta})$$

- [2,2]-Padé approximant

$$L_{\text{Padé}}(\beta) = \frac{a_0 + a_1(\beta - \bar{\beta}) + a_2(\beta - \bar{\beta})^2}{1 + b_1(\beta - \bar{\beta}) + b_2(\beta - \bar{\beta})^2}$$

Example on rare adverse events revisited



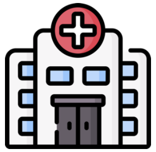
Fed-Padé algorithm

$\bar{\theta}$

Site 1

5 numbers: $g_1^0 = L_1(\bar{\theta}), g_1^1 = L_1^{(1)}(\bar{\theta}), \ldots, g_1^4 = L_1^{(4)}(\bar{\theta})$

Site 2



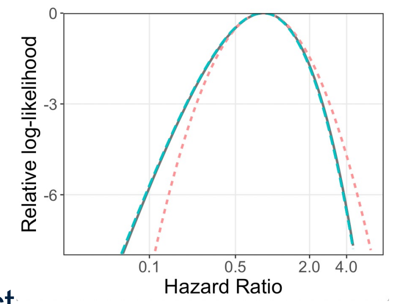
5 numbers: $g_2^0 = L_2(\bar{\theta}), g_2^1 = L_2^{(1)}(\bar{\theta}), \ldots, g_2^4 = L_2^{(4)}(\bar{\theta})$

⋮

Site K

5 numbers: $g_K^0 = L_K(\bar{\theta}), g_K^1 = L_K^{(1)}(\bar{\theta}), \ldots, g_K^4 = L_K^{(4)}(\bar{\theta})$

In a master site

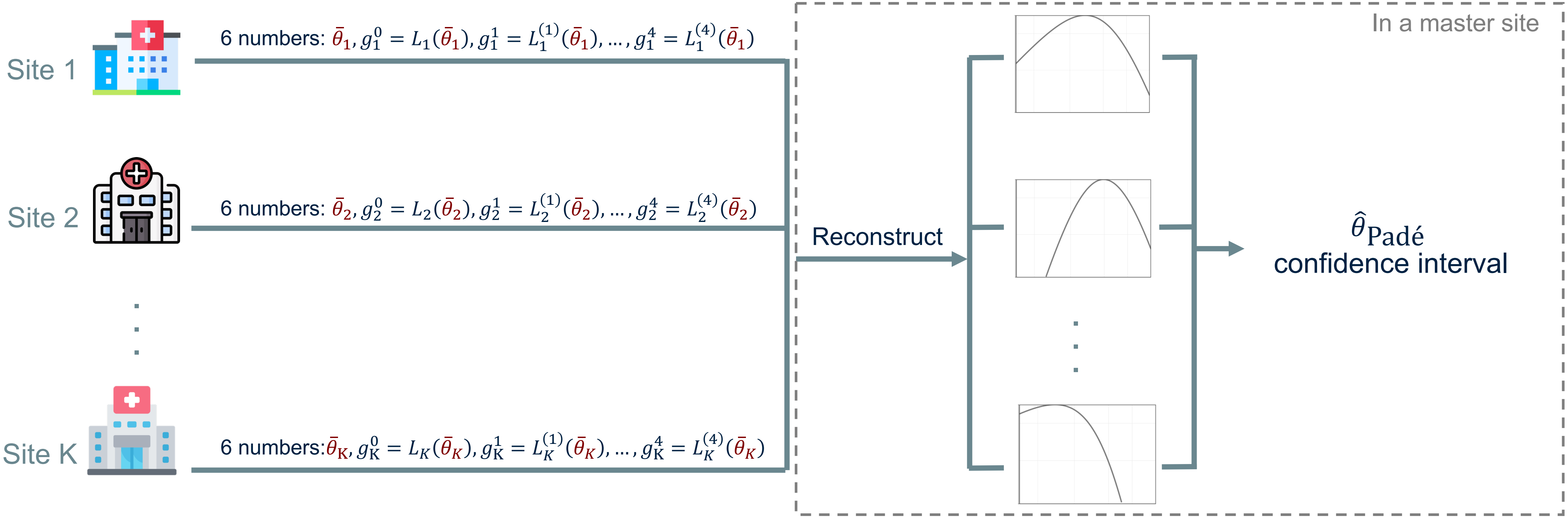


Reconstruct → $L_{\text{Padé}}(\theta)$ → $\hat{\theta}_{\text{Padé}}$

confidence interval

Random effects setting

- Effect sizes are i.i.d from normal distribution $\theta_j \sim N(\theta, \tau^2)$
- The normal approximation for per-site likelihood is problematic in rare events setting
  - Use Padé-approximated per-site likelihood instead of normal approximation



Stijnen, T., Hamza, T. H., & Özdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. Statistics in medicine, 29(29), 3046-3067.

Thank you for your time!

Journal of Biomedical Informatics 145 (2023) 104476

ELSEVIER

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Check for updates

Padé approximant meets federated learning: A nearly lossless, one-shot algorithm for evidence synthesis in distributed research networks with rare outcomes

Qiong Wu [a], Martijn J. Schuemie [b,c,d], Marc A. Suchard [b,d,e], Patrick Ryan [b,c,f], George M. Hripcsak [f,g], Charles A. Rohde [h], Yong Chen [a,b,*]

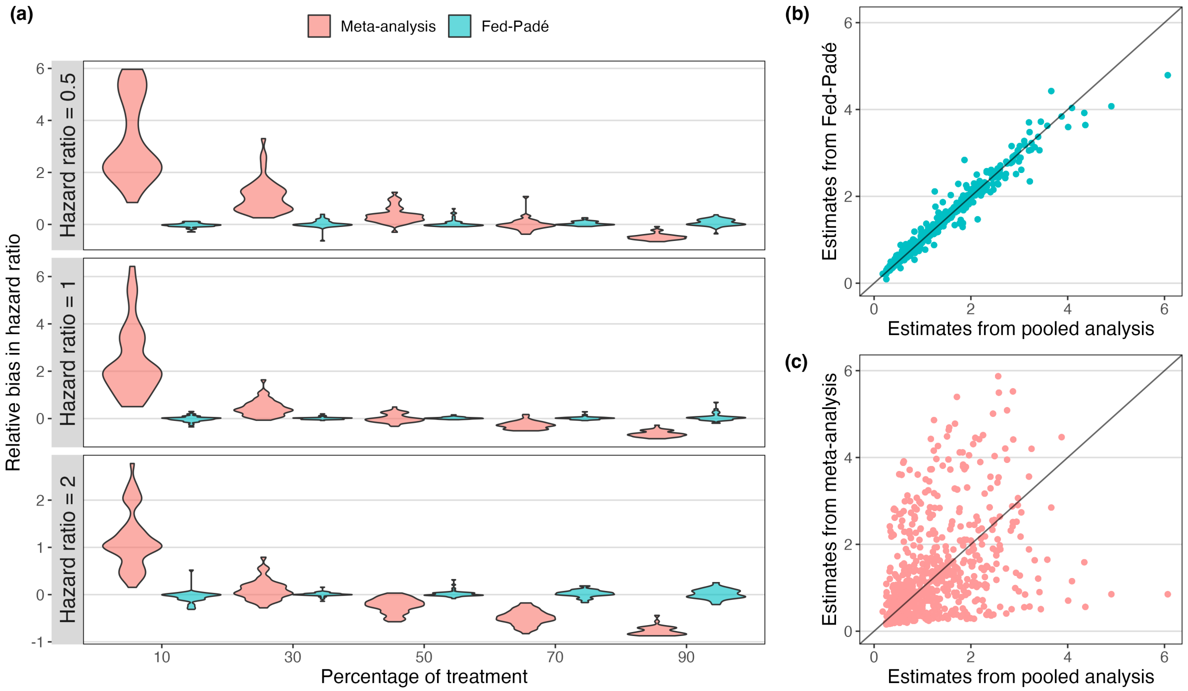
[a] Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States of America
[b] Observational Health Data Sciences and Informatics, New York, NY, United States of America
[c] Janssen Research & Development, Titusville, NJ, United States of America
[d] Department of Biostatistics, University of California, Los Angeles, CA, United States of America
[e] Department of Human Genetics, University of California, Los Angeles, CA, United States of America
[f] Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, United States of America
[g] Medical Informatics Services, New York-Presbyterian Hospital, New York, NY, United States of America
[h] Department of Biostatistics, Johns Hopkins University, Baltimore, MD, United States of America

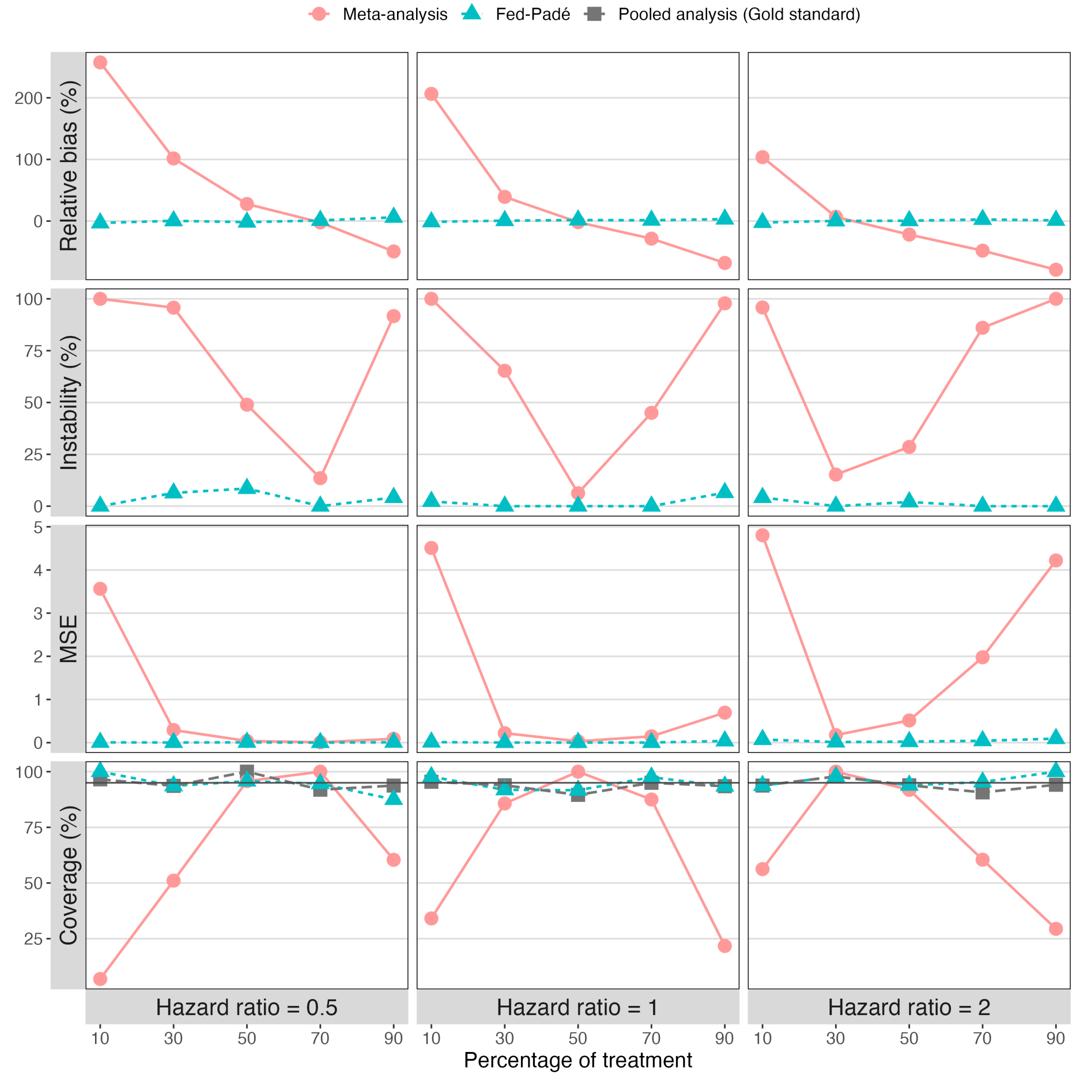
PDA

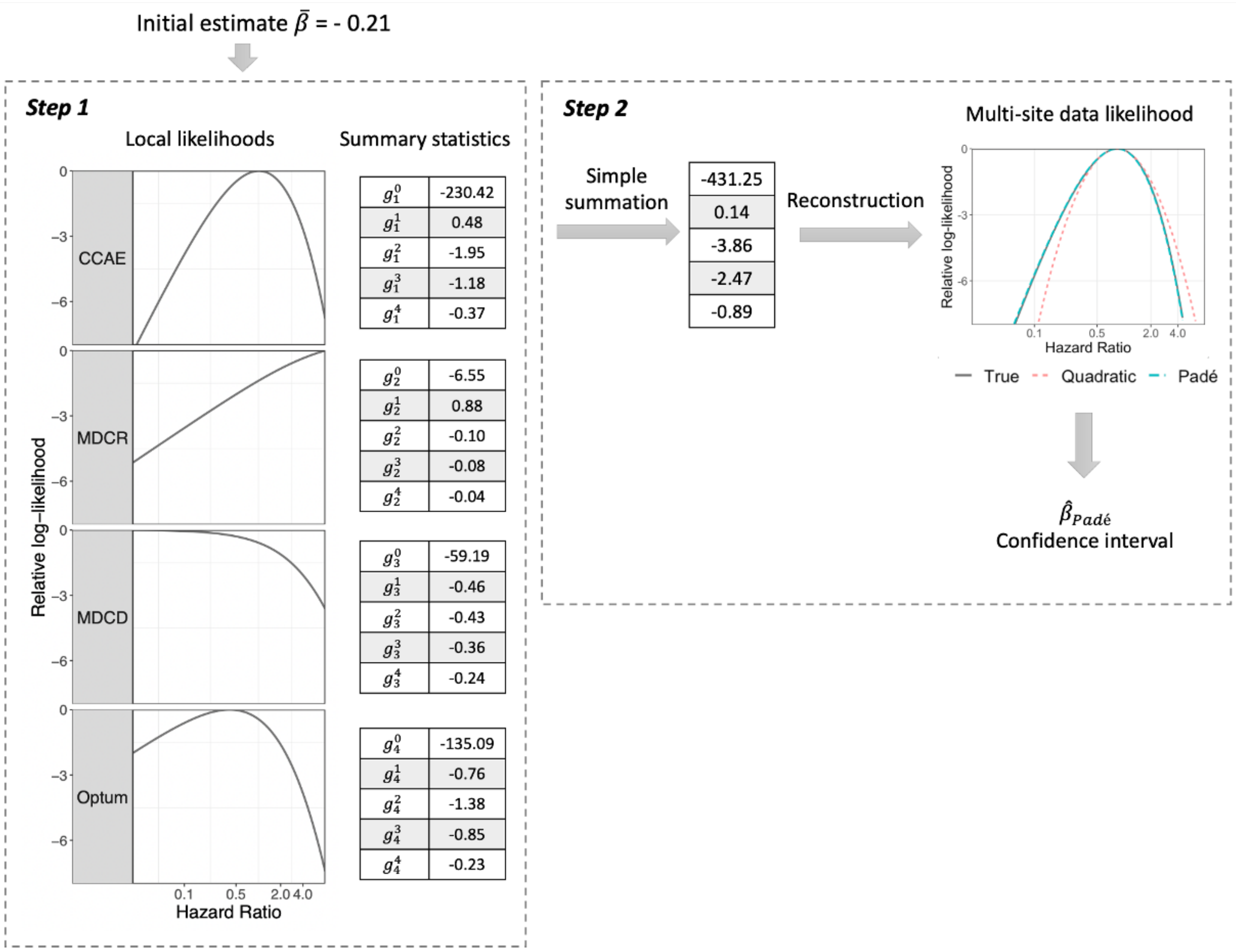
Privacy-preserving Distributed Algorithms

PDA-OTA

Simulation studies under random effect setting







Initial estimate $\bar{\beta}$ = - 0.21
Step 1
Local likelihoods
Summary statistics
Relative log−likelihood
CCAE
MDCR
MDCD
Optum
0
−3
−6
0.1
0.5
2.0 4.0
Hazard Ratio
g_1^0 -230.42
g_1^1 0.48
g_1^2 -1.95
g_1^3 -1.18
g_1^4 -0.37
g_2^0 -6.55
g_2^1 0.88
g_2^2 -0.10
g_2^3 -0.08
g_2^4 -0.04
g_3^0 -59.19
g_3^1 -0.46
g_3^2 -0.43
g_3^3 -0.36
g_3^4 -0.24
g_4^0 -135.09
g_4^1 -0.76
g_4^2 -1.38
g_4^3 -0.85
g_4^4 -0.23
Step 2
Simple summation
-431.25
0.14
-3.86
-2.47
-0.89
Reconstruction
Multi-site data likelihood
Relative log-likelihood
Hazard Ratio
True
Quadratic
Padé
$\hat{\beta}_{Padé}$
Confidence interval