Secure Federated Hospital Profiling via dGEM Algorithm using Real-World Data

Jiayi Tong[a], Jenna Reps[b,c], Chongliang Luo[d], Jiang Bian[e], Milou Brand[f], Zhaoyi Chen[e], Scott DuVall[g], Thomas Falconer[h], Kevin He[i], Miguel Angel Mayer[j], Bhavnisha Patel[g], Di Wang[i], Hua Xu[k], Yujia Zhou[k], Haitao Chu[l], Martijn Schuemie[b,c], Patrick Ryan[b,c], Rachel M.Werner[m,n], David A. Asch[m,n], Yong Chen[a,n]

a.Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

b. Janssen Research and Development, Titusville, NJ, USA

c. Observational Health Data Sciences and Informatics (OHDSI), New York, NY

d. Division of Public Health Sciences, Department of Surgery, Washington University in St. Louis, St. Louis, MO, USA

e. College of Medicine, Health Outcomes & Biomedical Informatics, University of Florida, Gainesville, FL, USA

f. IQVIA, England, United Kingdom

g. George E. Whalen Department of Veterans Affairs Health Care System, 500 Foothill Drive, Salt Lake City, UT, USA

h. Department of Biomedical Informatics, Columbia University, New York, NY, USA

i. Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

j. Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain

k. School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, Texas, USA

l. Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN

m. Division of General Internal Medicine, University of Pennsylvania, Philadelphia, PA, USA

n. Leonard Davis Institute of Health Economics, Philadelphia, PA, USA

Background

The real-world data (RWD), such as electronic health records data (EHRs), claims, and billing data among others, have become an invaluable data source for comparative effectiveness research (CER) during the past few years[1,2]. In comparison to the limited available data from a single site, the RWD synthesized from multiple clinical sites provides a larger sample size of the population for analysis[3]. The integration of research networks inside healthcare systems also allows rapid translation and dissemination of research findings into evidence-based healthcare decision-making to improve health outcomes. Several successful networks have become beneficial to multicenter research, such as the Observational Health Data Sciences and Informatics (OHDSI) consortium[4], the National Pediatric Learning Health System (PEDSnet) [5,6], Sentinel System[7], and the Consortium for Clinical Characterization of COVID-19 by EHR (4CE)[8], among others.

Beyond the improved accuracy in estimation, the study on hospital profiling, which is not possible within a single site, is also fascinating using multi-site data. Hospital profiling evaluates how much patient outcomes are influenced by the hospitals. Hospital profiling allows for a quantitative comparison of healthcare providers' quality of care for certain clinical outcomes, that guide the patients to choose the hospital to visit. In a recent article on hospital profiling for COVID-19 mortality[9], Asch et al. ranked the performance of 929 hospitals after adjusting for the patients' characteristics including age, sex, Elixhauser comorbidities, insurance type, and hospital characteristics including number of beds, number of ICU beds, urban/ nonurban setting,

geographic region, profit status, and academic affiliation. Research of this kind helps untangle what are often separate contributors to the production of good patient outcomes and is essential for identifying ways to improve those outcomes.

Toward the investigation of hospital profiling, if assuming individual-level data can be pooled together, the generalized linear mixed effects model (GLMM) is typically used to model the hospital-specific and patient-level effects simultaneously. However, there are a few challenges in hospital profiling using multi-site data in practice. The first challenge is that, for most the cases, the individual-level data from multiple institutions, databases, or networks are difficult to be centralized due to privacy considerations. The second challenge to be considered in developing the decentralized algorithm is the communication cost. Third, the data heterogeneity across multiple sites should be taken into account for in the hospital profiling.

We developed a novel decentralized algorithm for a generalized linear mixed effects model. We term the algorithm as dGEM. Specifically, the dGEM algorithm is expected to require aggregated data (AD) from each hospital in a single communication and obtain accurate estimates of the model parameters, and therefore accurate estimates of risk standardized event rates (RSER) for hospital profiling. We demonstrated the applicability of the proposed dGEM algorithm by hospital profiling using two real-world datasets, the kidney transplant registry data from 201 transplant centers and 12 international sites within the OHDSI network.

Methods

Let K denote the number of total clinical sites. n_k represents the number of patients in each clinical site, where the index of site $k \in \{1, \dots, K\}$. Let N denote the number of patients across K sites (i.e., $N = \sum_{k=1}^{K} n_k$). For the j-th patient from the k-th site, the response variable is denoted as y_{kj}, where the index of patient $j \in \{1, \dots, n_k\}$; the p-dimensional patient-level covariates is denoted as $\boldsymbol{x}_{kj}$ with fixed effect $\boldsymbol{\beta}$; the q-dimensional hospital-level covariates is denoted as $\boldsymbol{z_k}$ with random slope $\boldsymbol{\eta_k}$. The random intercept of the k-th site is denoted as γ_k. If the data can be pooled together across site, a pooled GLMM which is written as:

$$y_{kj} = h(\gamma_k + \boldsymbol{z_k^T \eta_k} + \boldsymbol{x}_{kj}{}^T \boldsymbol{\beta} + \epsilon_{kj})$$

where $\epsilon_{kj} \sim N(0, 1)$, h is the inverse link function, can be fitted to account for hospital-specific effects. However, when the patient-level data are stored in a decentralized format and only aggregated data are allowed to be shared across hospitals.

Thus, we developed a one-shot dGEM algorithm for GLMM estimation in the case that the individual patient data (IPD) are distributed across multiple centers and direct transfer of the IPD is not allowed. There are three main steps in the dGEM algorithm for hospital profiling. In Step I, the estimation of fixed effect can be achieved by meta-analyzing the estimated effects from all sites (Figure 1). With the shared meta-estimate of the fixed effect, each site estimates the hospital-specific effect, which is denoted as γ_k^* (Step II). The hospital-specific effects obtained from all sites are then calibrated through meta-regression with hospital-level covariates. The calibrated hospital effects γ_k^{**} are then broadcasted to the hospitals to calculate the directly

standardized mortality rates in Step III (denoted as DS in Figure 1)[10]. The directly standardized event rate measures can be calculated in a distributed manner without sharing IPD. For example, the DS for the k-th center can be obtained with the following equation:

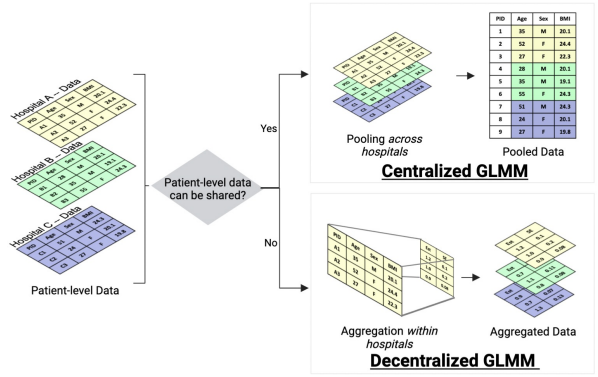
$$P_{k^*}^{DS} = \frac{\sum_{k=1}^{K} p_{k^*k}}{\sum_{k=1}^{K} n_k}$$

where

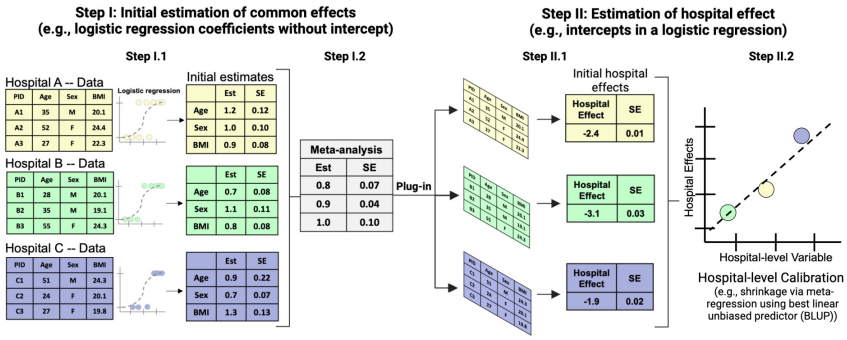
$$p_{k^*k}\left(\boldsymbol{X}_{k^*j}\right) = \text{expit}\left(\gamma_k^{**} + \boldsymbol{X}_{k^*j}^T \boldsymbol{\beta}\right).$$

and k represents the hypothetical attending hospital which contributes the estimated random intercept and k^* denotes the hospital that patient truly attended.

a. Feasibility of sharing patient-level data across hospitals



b. Steps of conducting decentralized GLMM (dGEM) based on aggregated data



c. Using results from decentralized GLMM (dGEM) to perform event rates calculation and hospital profiling

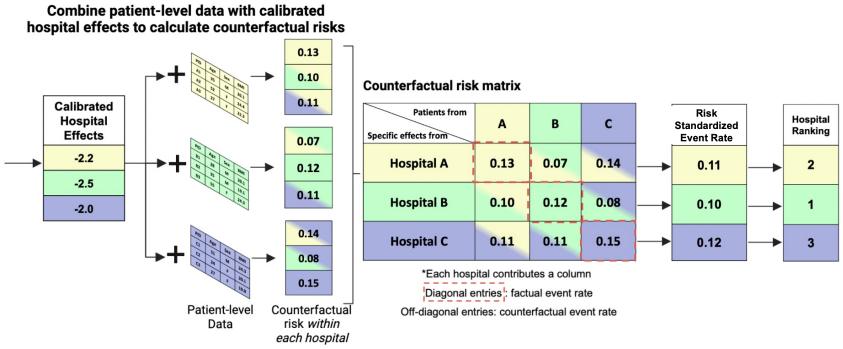


Fig. 1: Motivation and workflow of dGEM. a. Feasibility of sharing patient-level data across hospitals. Decentralized GLMM requiring aggregated data is needed when the patient-level data cannot be shared across sites. **b. Steps of conducting decentralized GLMM (dGEM) based on aggregated data** .In Step I, the estimated common effects of covariates across hospitals (Step I.1) are combined with the meta-analysis method (Step I.2) and transferred to all hospitals. In Step II, each site estimates the initial hospital-specific effects by plugging in the estimated common effect (Step II.1), followed with a meta-regression model to calibrate the hospital effects with hospital-level characteristics (Step II.2). **c. Using results from dGEM to perform event rates calculation and hospital profiling.** The risk standardized event rates (i.e., counterfactual event rates) are calculated using the estimated effects from previous steps. These measures are then compared across hospitals for hospital profiling.

Commented [H1]: Very good demonstration!

Commented [H2]: I suggest to split Step I and II each into two Steps, and call hospital effects after Step I as initial hospital effects.

Commented [TJ3R2]: This is a good idea. Will update this figure accordingly.

Results

Use case 1: The following **Figure 2** summarizes the results using the kidney transplant registry data collected from the U.S. Organ Procurement and Transplantation Network (OPTN). The data contain 44,428 adult deceased donor recipients who experienced a kidney transplant between January 1, 2008 and December 31, 2012 from 187 transplant centers. Given the centralized data, we demonstrated our proposed dGEM method by comparing it with fitting the GLMM on the centralized/pooled data. From Figure 2, we observed that the hospital rankings of the 201 centers with the proposed dGEM method and pooled GLMM method are close with a Kendall rank correlation value of 0.89 (p-value < 0.001). The results demonstrate that the proposed dGEM method has a similar performance as the pooled GLMM in hospital profiling.

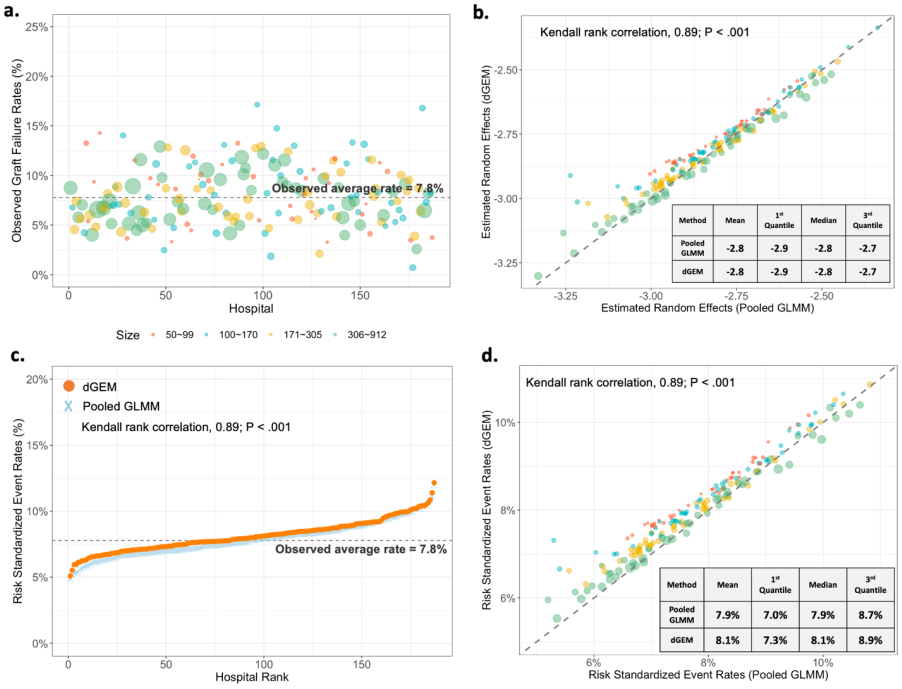


Fig. 2: Validation with *centralized* kidney transplant data. a. Summary of 187 transplant centers data. Each circle represents one center, and the size of the circle is proportional to the sample size. The colors of the circles represent the size groups. **b.** Comparison between the estimated random effects with pooled GLMM (x-axis) and the proposed dGEM framework (y-axis). **c.** Comparison between the hospital ranking based on the risk standardized event rates (i.e., kidney graft failure rates) of the two methods. A numerically higher rank corresponds to worse performance. **d.** Comparison between the risk standardized event rates (RSER, i.e., counterfactual event rates) with pooled GLMM (x-axis) and the proposed dGEM framework (y-axis).

Commented [H4]: Why directly? It seems that dGEM overestimate the SER compared to pooled GLMM? Bland-Alman plot for b and d? For c, should we use the difference between dGEM and pooled GLMM? Are we able to present 95% CI of the difference?

Use case 2: We demonstrated the proposed dGEM framework by utilizing it in a truly decentralized real-world setting. We aimed to study the variation in hospital profiling of COVID-19 mortality rate within the international Observational Health Data Sciences and Informatics (OHDSI). The OHDSI network has accumulated more than half a billion patient records from 19 different countries worldwide, with around 300 million patient records within the United States21. This case study was termed dGEM-COVID.

The dGEM-COVID study started with recruitment and protocol development in May 2022. By June 2022, 12 sites from 4 countries (the US, Spain, Netherlands, and the UK) joined the study.

To examine the temporal changes in hospital profiling, we analyzed two time periods. We termed the first period Wave A, which is around October 2020 to April 2021. Though the major variant that attacked the whole world during this period was the Alpha variant, the existence of other variants and the different variant surges that occurred in different countries led us to name the period "Wave A" instead of "Alpha Wave" or "Wave 1" to avoid confusion. Similarly, we termed the second period Wave B, which represents the period from May 2021 to October 2021. Since different countries or areas would be hit at different times, we used a data-driven method to control the data quality, thereby determining the timeframe of the waves for each dataset. More details are provided in the Methods section.

The study cohort consisted of patients ages 18 or older who had an inpatient visit with a diagnosis of COVID-19 a positive test for COVID-19 occurring between 21 days prior to the visit and the end of the visit. Patients were excluded if they had been observed in the database for less than 180 days. The sample sizes varied greatly across 12 sites, ranging from 100 to 81,530. The patient's disease history, Charlson index, and age distributions were heterogeneous as well. Based on Fig 2d and 2e, we observed that Sites 2,3,5,6 had more severe (larger proportion of history of conditions and Charlson index larger than 5) and older patients. The outcome of each patient is the binary mortality status indicating whether the patient died during the inpatient visit or within 7 days after the visit and the covariates considered are summarized in Table 1 in the Methods section.

In terms of implementing the dGEM framework, a web-based secure platform PDA-OTA (Privacy-preserving Distributed Algorithm Over the Air, https://pda-ota.pdamethods.org/home/) was employed, which enables synchronization of project information and status, allocation of aggregated data (AD), and encryption of hospital information. This platform served as the coordinating center for the collaborating sites to upload and manage the AD. In this dGEM-COVID study, there were a total of three rounds of communications. Each site is responsible to accomplish two tasks within each round: download the control file from PDA-OTA and execute an R study package to generate AD with local patient-level data to upload to the PDA-OTA platform. In this implementation, though each step took about one month to accomplish, the efforts due to human-in-the-loop required more time in practice than that by the computational time of performing the framework. Within each step for one wave, a single site (e.g., Janssen) only needs less than 1 hour to implement the required steps, including downloading control file (i.e., summary of study information, such as statistical model, outcome, covariates, etc), executing the R code to produce AD, and uploading AD to the coordinating platform, PDA-OTA.

Figure 3 presents the results of the dGEM-COVID study.

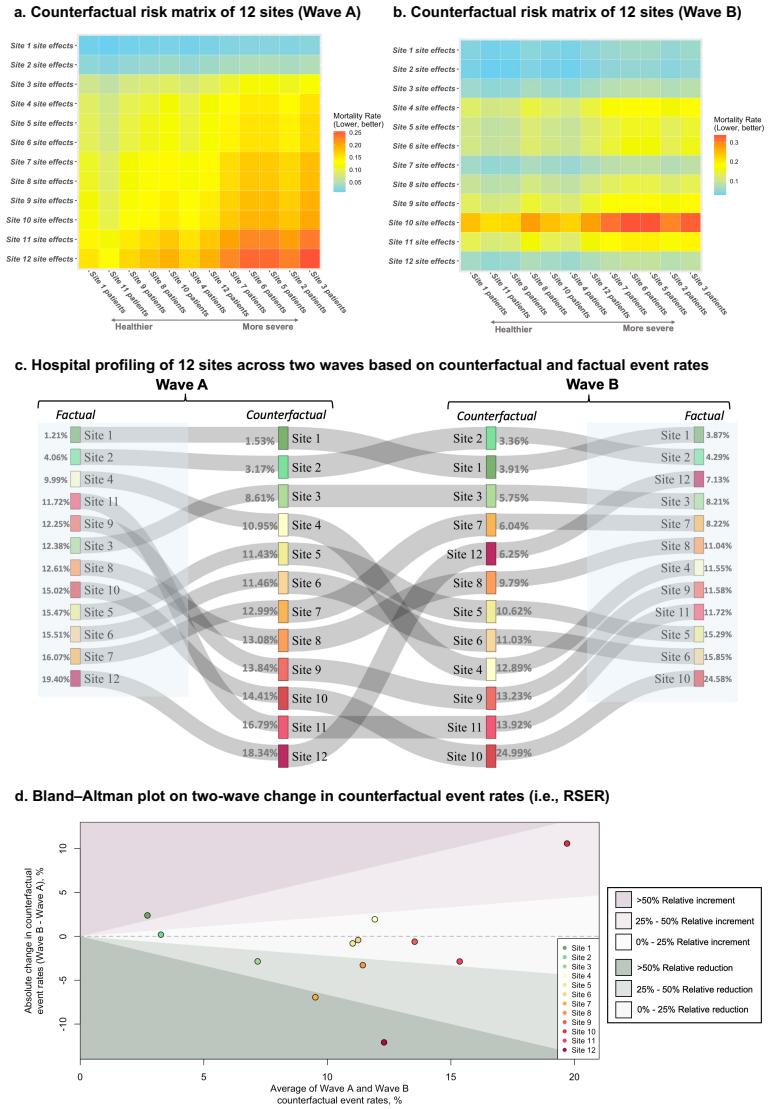


Fig. 3: Application to COVID-19 data within OHDSI network. a. Counterfactual risk matrix of 12 sites (Wave A). Each column represents the group of patients from a specific site, and each row represents the hospital-specific effect. On x-axis, the sites are ranked by patient severity (i.e., proportion of patients with Charlson

index larger than 5) b. Counterfactual risk matrix of 12 sites (Wave B). c. Hospital profiling of 12 sites across two waves based on counterfactual and factual event rates. d. Bland–Altman plot on two-wave change in counterfactual event rates (i.e., RSER)

Conclusion

The dGEM algorithm is an effective distributed/federated learning algorithm for GLMM to study hospital profiling. dGEM provides accurate estimated fixed and random effects while accounting for heterogeneity and hospital-level calibration. We are currently illustrating the applicability of dGEM in an OHDSI study with multiple international hospitals where the patient-level data cannot be shared.

References/Citations

1. Sherman, R. E. *et al.* Real-world evidence—what is it and what can it tell us. *N. Engl. J. Med.* **375**, 2293–2297 (2016).
2. Fda, U. S., Food, Administration, D. & Others. Framework for FDA's Real-World Evidence Program. (2018).
3. Bowens, F. M., Frye, P. A. & Jones, W. A. Health information technology: integration of clinical workflow into meaningful use of electronic health records. *Perspect. Health Inf. Manag.* **7**, 1d (2010).
4. Hripcsak, G. *et al.* Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud. Health Technol. Inform.* **216**, 574–578 (2015).
5. Forrest, C. B. *et al.* PEDSnet: a National Pediatric Learning Health System. *J. Am. Med. Inform. Assoc.* **21**, 602–606 (2014).
6. Fleurence, R. L. *et al.* Launching PCORnet, a national patient-centered clinical research network. *J. Am. Med. Inform. Assoc.* **21**, 578–582 (2014).
7. Platt, R. *et al.* The US Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiol. Drug Saf.* **21**, 1–8 (2012).
8. Brat, G. A. *et al.* International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. *Npj Digital Medicine* **3**, 1–9 (2020).
9. Asch, D. A. *et al.* Variation in US hospital mortality rates for patients admitted with COVID-19 during the first 6 months of the pandemic. *JAMA Intern Med* **181**, 471–478 (2021).
10. George, E. I., Ročková, V., Rosenbaum, P. R., Satopää, V. A. & Silber, J. H. Mortality rate estimation and standardization for public reporting: Medicare's hospital compare. *J Am Stat Assoc* **112**, 933–947 (2017).