

iScience, Volume 27

Supplemental information

Federated difference-in-differences with multiple time periods in DataSHIELD

Manuel Huth, Carolina Alvarez Garavito, Lea Seep, Laia Cirera, Francisco Saúte, Elisa Sicuri, and Jan Hasenauer

Summary statistics - Simulated dataset

Table S1: Summary statistics: Number of individuals for the simulated dataset. Related to Figure 3. N indicates the total number of individuals at the respective server or the pooled data. The columns below *Treated in* indicate the periods in which individuals have received treatment. No individual was treated in period 1 as they could not serve as a comparison group.

	N	Never Treated	Treated in 2	Treated in 3	Treated in 4
Pooled	801	222	168	195	216
Server 1	134	32	23	36	43
Server 2	134	38	27	39	30
Server 3	134	35	30	31	38
Server 4	133	38	28	30	37
Server 5	133	33	33	29	38
Server 6	133	46	27	30	30

Summary statistics - Malaria dataset

Table S2: Summary statistics: Average performance of students by district, school and term of studies. Related to Figure 4. Standard deviations are shown in parenthesis for all sample averages. For number of females, the table reports the percentage of female students on the total school population N.

					2015	2015	2015	2016	2016	2016	
	District	School	N	Mean age	Females	Term 1	Term 2	Term 3	Term 1	Term 2	Term 3
1	Magude	Duco	28	9.36	14	12.26	12.12	12.20	12.68	12.41	12.54
				(1.56)	(50%)	(2.07)	(2.22)	(2.35)	(2.21)	(2.37)	(2.39)
2	Magude	Graca Machel	90	9.43	55	12.70	12.53	12.33	12.73	12.90	12.89
				(2.59)	(61%)	(2.16)	(2.53)	(2.57)	(1.87)	(1.93)	(1.92)
3	Magude	Maguiguana	89	9.33	49	12.09	11.86	11.86	12.41	12.38	12.59
				(1.83)	(55%)	(2.09)	(2.23)	(2.36)	(2.32)	(2.56)	(2.78)
4	Magude	Moine	21	10.52	9	11.36	11.67	11.23	12.43	13.01	12.86
				(1.60)	(43%)	(1.26)	(1.49)	(2.09)	(1.70)	(2.14)	(2.60)
5	Magude	Simbe	22	9.68	8	11.97	11.53	11.32	11.50	11.44	11.13
				(2.90)	(36%)	(1.85)	(1.69)	(1.60)	(1.68)	(1.71)	(1.64)
6	Manhica	3 de Fev	208	9.90	90	11.75	11.71	11.76	11.79	11.81	11.88
				(2.13)	(43%)	(2.03)	(2.13)	(2.29)	(2.01)	(2.13)	(2.31)
7	Manhica	Ilha Josina	156	10.04	79	12.81	12.91	12.70	12.77	12.65	12.58
				(2.36)	(51%)	(1.64)	(1.60)	(1.65)	(1.47)	(1.54)	(1.54)
8	Manhica	Maragra	304	9.24	136	12.03	12.09	12.15	12.08	12.41	12.37
				(1.88)	(45%)	(1.91)	(2.17)	(2.07)	(2.02)	(2.00)	(1.96)
9	Manhica	Xinavane	126	9.88	61	13.46	13.21	12.73	13.03	13.24	12.96
				(2.74)	(48%)	(2.04)	(2.34)	(2.49)	(2.42)	(2.33)	(2.41)

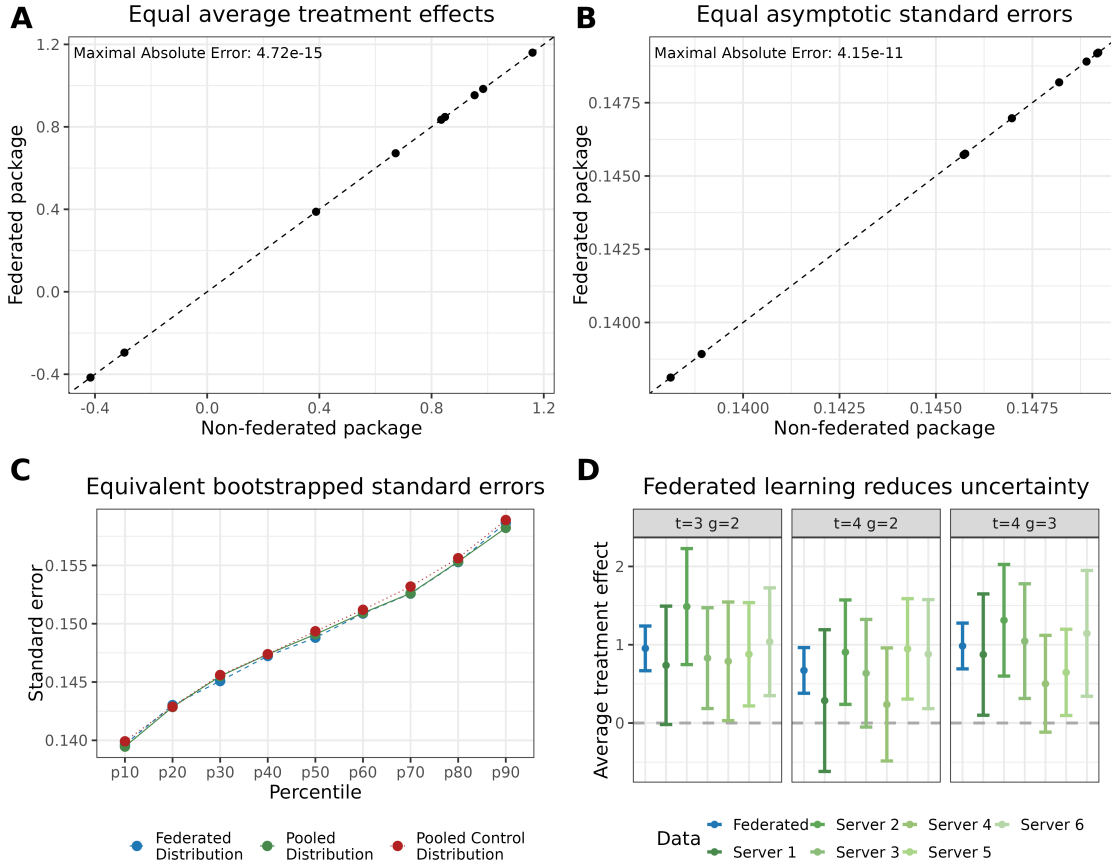


Figure S1: DR estimate and never treated individuals as control group. Related to Figure 3. The federated set-up consisted of 6 servers. Three servers contained 134 (536) individuals (observations), and three servers contained 133 (532) individuals (observations). Individuals were either never treated or treated in period two or three. All observations of one individual were within one server. **A** Depicts the equality of central and federated point estimates. The x-axis represents the point estimates of the central (non-federated) estimator, while the y-axis represents the estimates obtained from our federated approach. The diagonal line depicts the 45° line, indicating that the federated estimate yields results equivalent to the non-federated estimate when they align along this line. **B** Depicts the equality of central and federated asymptotic standard errors. The x-axis represents the asymptotic standard errors of the central (non-federated) estimator, while the y-axis represents the asymptotic standard errors obtained from our federated approach. The diagonal line again depicts the 45° line. **C** The subplot displays a comparison of the distribution of bootstrapped standard errors using percentiles of the distribution function. The distribution of the federated bootstrapped standard errors is shown in blue, while the distribution of the central estimation is shown in green (baseline) and red (control). Two central learning distributions were computed to establish a plausible reference difference between two equal distributions. A total of 500 bootstrapped standard errors were computed to obtain the distributions for analysis. **D** The subplot presents point estimates (depicted as dots) and 95% confidence intervals of the estimated treatment effects for the periods following the treatment. The estimates obtained with our federated package are shown in blue, while estimates based on data from only one server are depicted in green. This figure has been designed using resources from Flaticon.com

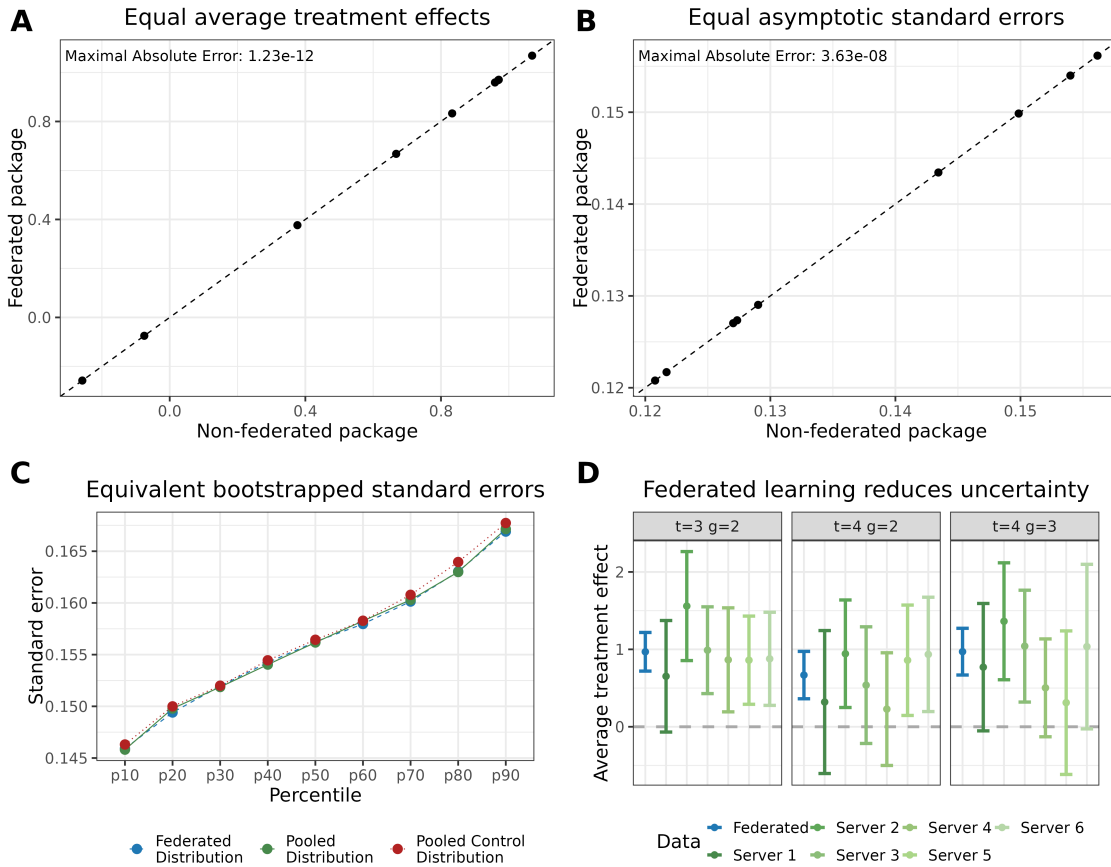


Figure S2: Inverse probability weighted estimate and not-yet treated individuals as control group. Related to Figure 3. The federated set-up consisted of 6 servers. Three servers contained 134 (536) individuals (observations), and three servers contained 133 (532) individuals (observations). Individuals were either never treated or treated in period two or three. All observations of one individual were within one server. **A** Depicts the equality of central and federated point estimates. The x-axis represents the point estimates of the central (non-federated) estimator, while the y-axis represents the estimates obtained from our federated approach. The diagonal line depicts the 45° line, indicating that the federated estimate yields results equivalent to the non-federated estimate when they align along this line. **B** Depicts the equality of central and federated asymptotic standard errors. The x-axis represents the asymptotic standard errors of the central (non-federated) estimator, while the y-axis represents the asymptotic standard errors obtained from our federated approach. The diagonal line again depicts the 45° line. **C** The subplot displays a comparison of the distribution of bootstrapped standard errors using percentiles of the distribution function. The distribution of the federated bootstrapped standard errors is shown in blue, while the distribution of the central estimation is shown in green (dark and light). Two central learning distributions were computed to establish a plausible reference difference between two equal distributions. A total of 500 bootstrapped standard errors were computed to obtain the distributions for analysis. **D** The subplot presents point estimates (depicted as dots) and 95% confidence intervals of the estimated treatment effects for the periods following the treatment. The estimates obtained with our federated package are shown in blue, while estimates based on data from only one server are depicted in green. This figure has been designed using resources from Flaticon.com

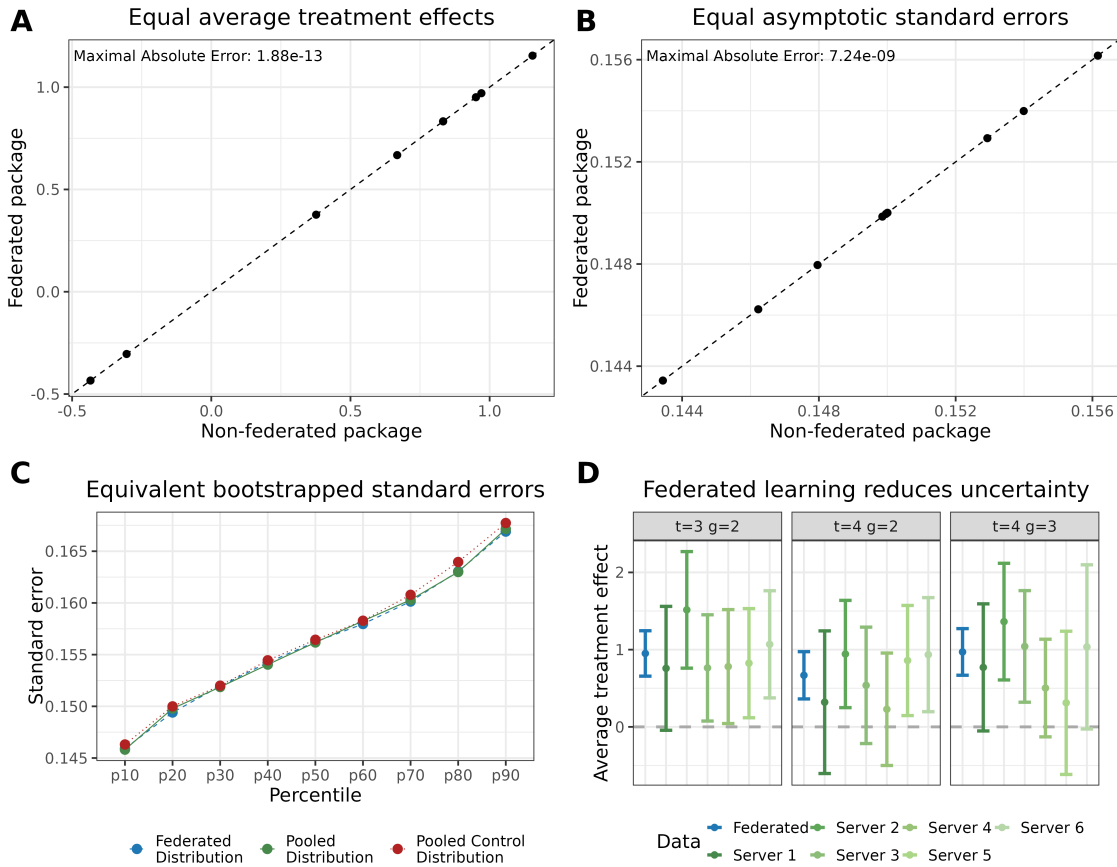


Figure S3: Inverse probability weighted estimate and never treated individuals as control group. Related to Figure 3. The federated set-up consisted of 6 servers. Three servers contained 134 (536) individuals (observations), and three servers contained 133 (532) individuals (observations). Individuals were either never treated or treated in period two or three. All observations of one individual were within one server. **A** Depicts the equality of central and federated point estimates. The x-axis represents the point estimates of the central (non-federated) estimator, while the y-axis represents the estimates obtained from our federated approach. The diagonal line depicts the 45° line, indicating that the federated estimate yields results equivalent to the non-federated estimate when they align along this line. **B** Depicts the equality of central and federated asymptotic standard errors. The x-axis represents the asymptotic standard errors of the central (non-federated) estimator, while the y-axis represents the asymptotic standard errors obtained from our federated approach. The diagonal line again depicts the 45° line. **C** The subplot displays a comparison of the distribution of bootstrapped standard errors using percentiles of the distribution function. The distribution of the federated bootstrapped standard errors is shown in blue, while the distribution of the central estimation is shown in green (dark and light). Two central learning distributions were computed to establish a plausible reference difference between two equal distributions. A total of 500 bootstrapped standard errors were computed to obtain the distributions for analysis. **D** The subplot presents point estimates (depicted as dots) and 95% confidence intervals of the estimated treatment effects for the periods following the treatment. The estimates obtained with our federated package are shown in blue, while estimates based on data from only one server are depicted in green. This figure has been designed using resources from Flaticon.com

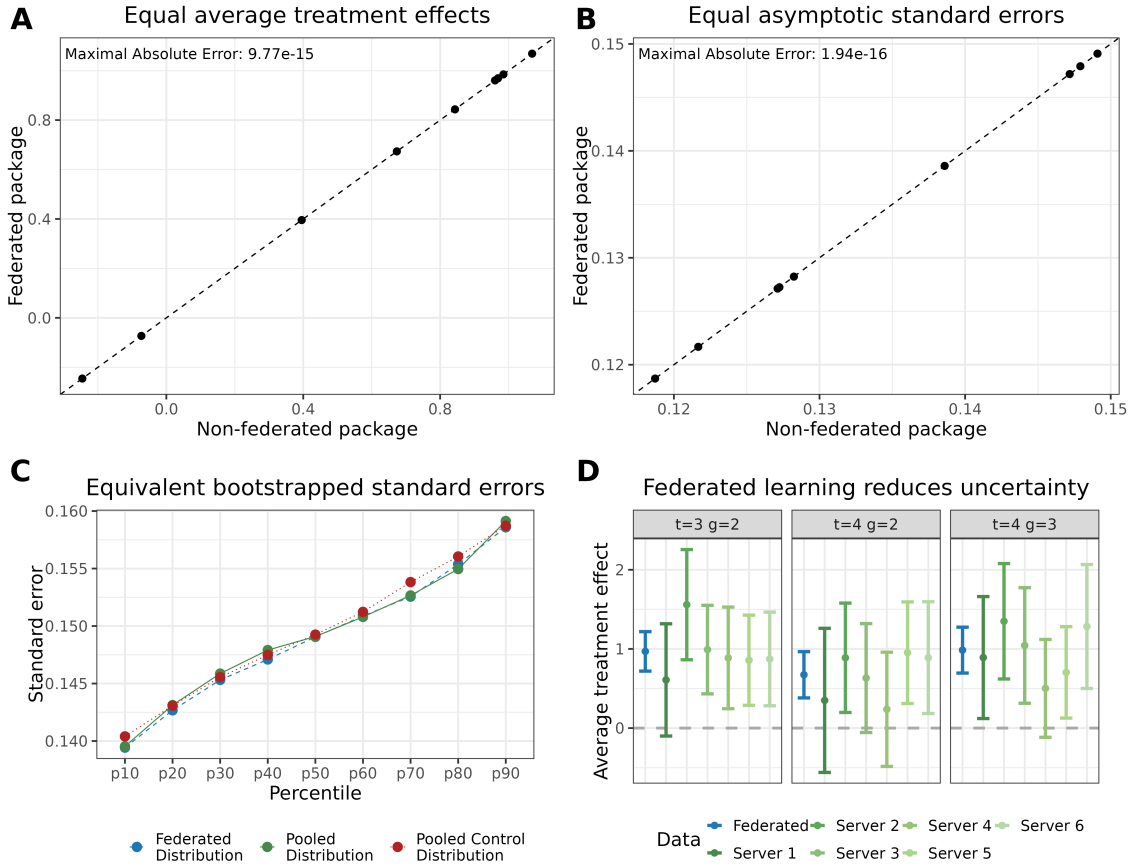


Figure S4: Outcome regression estimate and not-yet treated individuals as control group. Related to Figure 3. The federated set-up consisted of 6 servers. Three servers contained 134 (536) individuals (observations), and three servers contained 133 (532) individuals (observations). Individuals were either never treated or treated in period two or three. All observations of one individual were within one server. **A** Depicts the equality of central and federated point estimates. The x-axis represents the point estimates of the central (non-federated) estimator, while the y-axis represents the estimates obtained from our federated approach. The diagonal line depicts the 45° line, indicating that the federated estimate yields results equivalent to the non-federated estimate when they align along this line. **B** Depicts the equality of central and federated asymptotic standard errors. The x-axis represents the asymptotic standard errors of the central (non-federated) estimator, while the y-axis represents the asymptotic standard errors obtained from our federated approach. The diagonal line again depicts the 45° line. **C** The subplot displays a comparison of the distribution of bootstrapped standard errors using percentiles of the distribution function. The distribution of the federated bootstrapped standard errors is shown in blue, while the distribution of the central estimation is shown in green (dark and light). Two central learning distributions were computed to establish a plausible reference difference between two equal distributions. A total of 500 bootstrapped standard errors were computed to obtain the distributions for analysis. **D** The subplot presents point estimates (depicted as dots) and 95% confidence intervals of the estimated treatment effects for the periods following the treatment. The estimates obtained with our federated package are shown in blue, while estimates based on data from only one server are depicted in green. This figure has been designed using resources from Flaticon.com

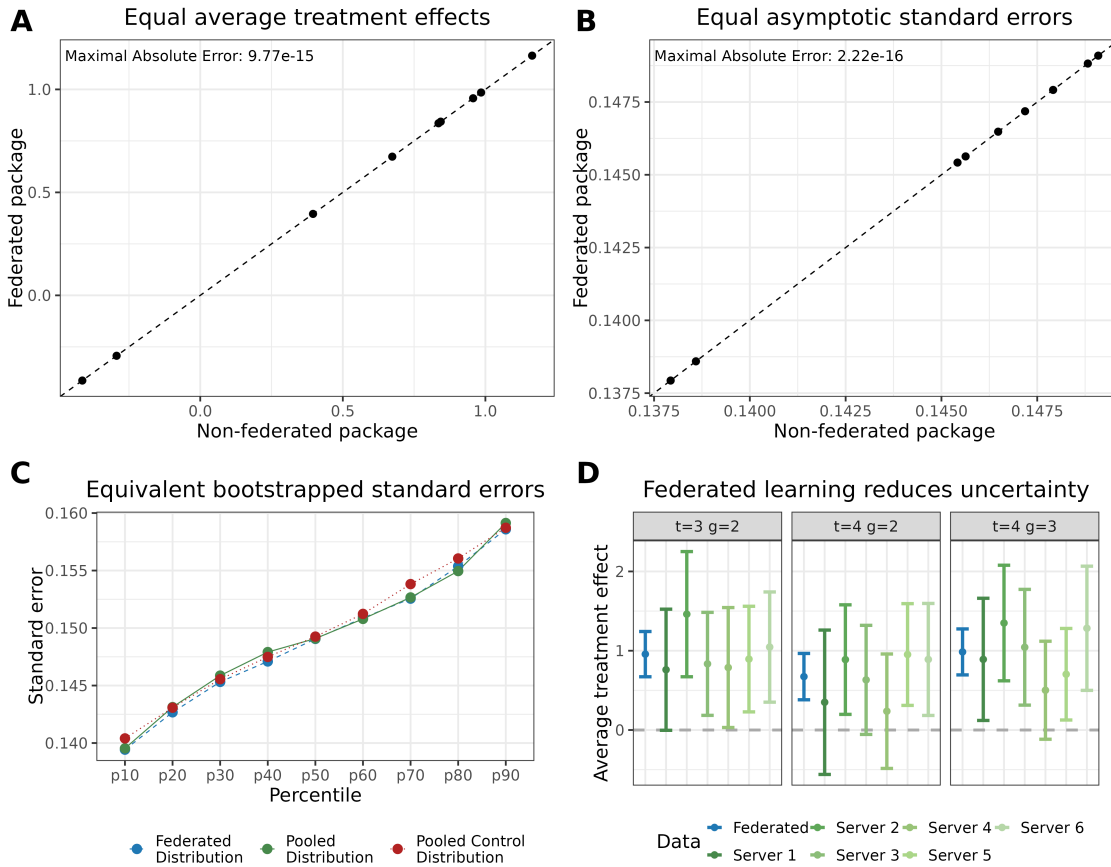


Figure S5: Outcome regression estimate and never treated individuals as control group. Related to Figure 3. The federated set-up consisted of 6 servers. Three servers contained 134 (536) individuals (observations), and three servers contained 133 (532) individuals (observations). Individuals were either never treated or treated in period two or three. All observations of one individual were within one server. **A** Depicts the equality of central and federated point estimates. The x-axis represents the point estimates of the central (non-federated) estimator, while the y-axis represents the estimates obtained from our federated approach. The diagonal line depicts the 45° line, indicating that the federated estimate yields results equivalent to the non-federated estimate when they align along this line. **B** Depicts the equality of central and federated asymptotic standard errors. The x-axis represents the asymptotic standard errors of the central (non-federated) estimator, while the y-axis represents the asymptotic standard errors obtained from our federated approach. The diagonal line again depicts the 45° line. **C** The subplot displays a comparison of the distribution of bootstrapped standard errors using percentiles of the distribution function. The distribution of the federated bootstrapped standard errors is shown in blue, while the distribution of the central estimation is shown in green (dark and light). Two central learning distributions were computed to establish a plausible reference difference between two equal distributions. A total of 500 bootstrapped standard errors were computed to obtain the distributions for analysis. **D** The subplot presents point estimates (depicted as dots) and 95% confidence intervals of the estimated treatment effects for the periods following the treatment. The estimates obtained with our federated package are shown in blue, while estimates based on data from only one server are depicted in green. This figure has been designed using resources from Flaticon.com

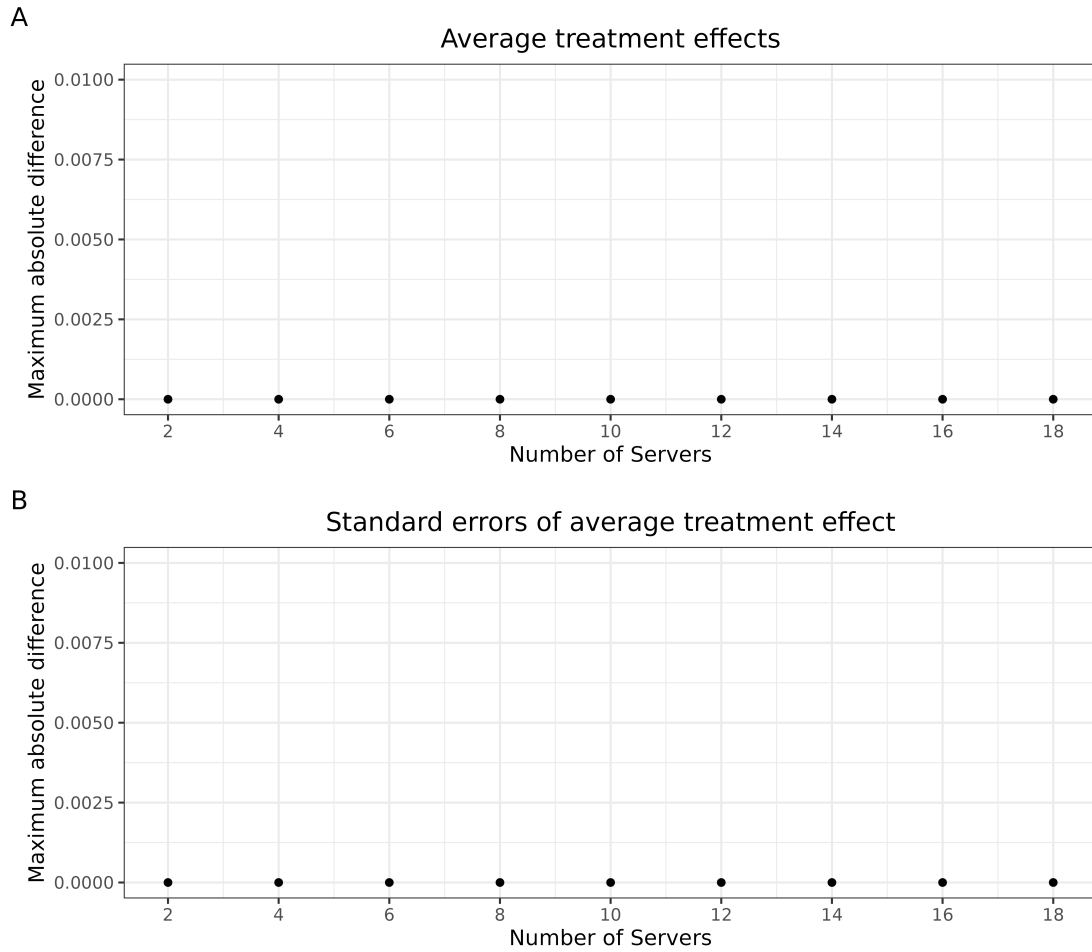


Figure S6: Differences between federated and non-federated estimates with respect to the number of servers. Related to Figure 3. Note that the sample size was held constant over the number of servers. **A** Depicts the equality of central and federated point estimates. **B** Depicts the equality of central and federated asymptotic standard errors.

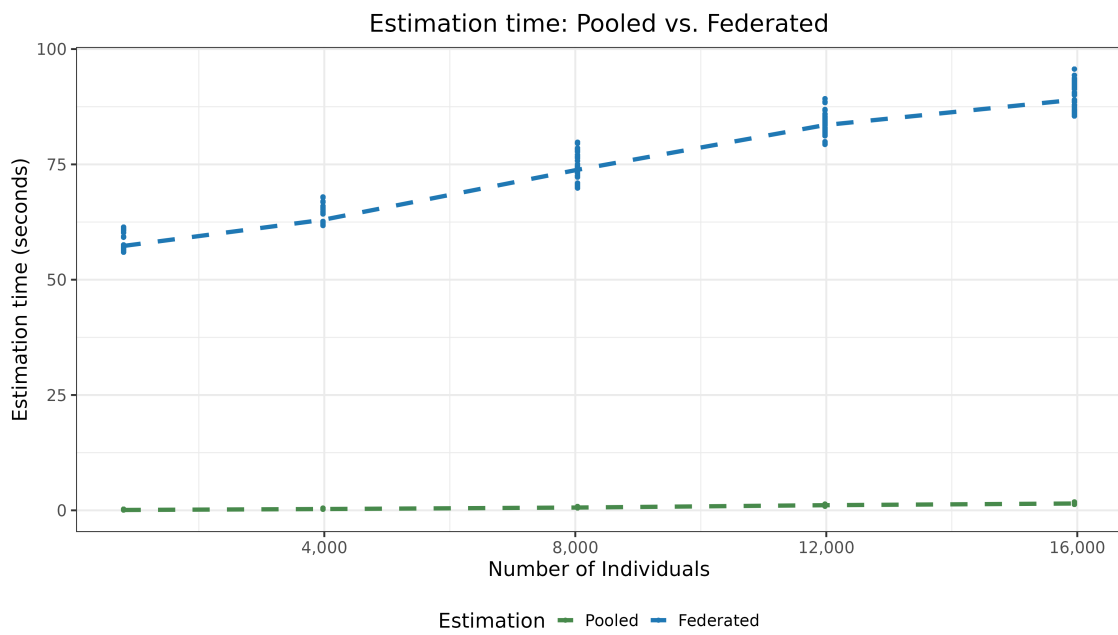


Figure S7: Comparison of estimation times between federated and non-federated (pooled) methods across varying numbers of individuals. Related to Figure 3. For each number of individuals, the estimation was performed 50 times using a doubly robust estimator with untreated individuals serving as the control group. The dashed lines connect the mean estimation times for each method.