Supplementary Appendix for the Simulation Studies in: "Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records."[*]

Ted Enamorado[†]   Benjamin Fifield[‡]   Kosuke Imai[§]

October 4, 2018

In this section we present the full set of simulation results. The tables present the estimation error for the turnout rate among matched voters, its corresponding root mean square error, the false discovery rate (FDR), the false negative rate (FNR), the estimated FDR and FNR, and the match rate of exact matching. The size of the datasets and the amount of overlap between them vary across each simulation setup. To declare a pair of records to be in the set of matches, three different thresholds are considered. The Online Supplementary Information Appendix S9 presents the details about how the data sets for these simulation exercises were constructed.

For the models assuming independence across linkage fields conditional on the matching status, two major are found under these other simulation scenarios. First, regardless of the missing data or measurement error mechanism structures, fastLink controls FDR, FNR, and estimation error well. Second, a greater degree of overlap between datasets leads to better merging results in terms of FDR and FNR as well as the accuracy of their estimates. Once we move away from the conditional independent model and dependencies are allowed, not only FDR, FNR, and the estimation error are controlled, but even in cases where the degree of overlap is small the parameters of the model are more accurate.

[†]Ph.D. Candidate, Department of Politics, Princeton University, Princeton NJ 08544. Email: `tede@princeton.edu`

[‡]Ph.D. Candidate, Department of Politics, Princeton University, Princeton NJ 08544. Email: `bfifield@princeton.edu`, URL: http://www.benfifield.com

[§]Professor, Department of Government and Statistics, Harvard University. Phone: 617–384–6778, Email: `imai@harvard.edu`, URL: `https://imai.fas.harvard.edu`

Simulation Studies Index

Scenario	Size Balance	Degree of Overlap	Conditional independence
Missing information under MAR	1:100	20% (Table S.A.1); 50% (Table S.A.2); 80% (Table S.A.3)	✓
	1:10	20% (Table S.A.4); 50% (Table S.A.5); 80% (Table S.A.6)	✓
	1:1	20% (Table S.A.7); 50% (Table S.A.8); 80% (Table S.A.9)	✓
Missing information as disagreements	1:100	20% (Table S.A.10); 50% (Table S.A.11); 80% (Table S.A.12)	✓
	1:10	20% (Table S.A.13); 50% (Table S.A.14); 80% (Table S.A.15)	✓
	1:1	20% (Table S.A.16); 50% (Table S.A.17); 80% (Table S.A.18)	✓
Measurement error	1:100	20% (Table S.A.19); 50% (Table S.A.20); 80% (Table S.A.21)	✓
	1:10	20% (Table S.A.22); 50% (Table S.A.23); 80% (Table S.A.24)	✓
	1:1	20% (Table S.A.25); 50% (Table S.A.26); 80% (Table S.A.27)	✓
Measurement error	1:100	20% (Table S.A.28); 50% (Table S.A.29); 80% (Table S.A.30)	✗
	1:10	20% (Table S.A.31); 50% (Table S.A.32); 80% (Table S.A.33)	✗
	1:1	20% (Table S.A.34); 50% (Table S.A.35); 80% (Table S.A.36)	✗

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-1.50	3.38	27.01	0.00	9.08	7.20	-1.10	3.35	20.00	0.00	1.73	40.65	-1.93	3.83	13.04	0.00	0.20	50.62	57.50
10%	0.50	3.21	21.60	2.00	1.29	21.73	-0.42	3.32	15.58	2.50	0.99	40.90	-1.50	3.68	12.61	3.00	0.18	45.55	40.50
15%	-1.04	3.56	13.89	7.00	3.09	31.68	-1.50	3.80	10.68	8.00	0.57	38.52	-2.01	4.09	8.08	9.00	0.19	41.08	28.00
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.53	3.09	27.68	2.00	1.51	9.58	-0.46	3.26	18.67	2.00	1.06	37.43	-1.28	3.55	14.41	2.00	0.18	44.85	60.50
10%	1.73	3.54	25.86	2.50	6.59	16.13	1.17	3.42	20.58	3.50	1.94	39.69	-0.17	3.34	14.60	3.50	0.40	50.00	47.50
15%	1.39	3.50	25.21	9.50	6.41	25.20	0.73	3.42	20.54	11.00	2.58	43.03	2.12	4.07	14.49	11.50	0.45	56.38	28.50
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	-0.60	3.07	28.52	1.00	2.26	11.88	-1.50	3.52	20.16	1.00	2.23	37.62	-1.07	3.45	15.81	1.50	0.24	50.32	59.50
10%	0.93	3.20	26.97	2.50	6.79	15.86	1.22	3.46	18.83	3.00	2.19	39.03	0.99	3.51	12.22	3.00	0.19	52.95	43.00
15%	-0.30	3.18	23.20	4.00	5.45	23.50	1.11	3.48	18.26	6.00	2.01	40.45	1.25	3.61	14.68	7.00	0.39	51.97	30.00
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-1.13	3.25	26.67	1.00	8.96	6.45	-0.87	3.37	16.88	1.50	1.38	41.76	-1.07	3.47	14.78	2.00	0.14	49.78	62.50
10%	-1.69	3.51	26.79	3.00	5.86	17.95	-0.88	3.33	20.58	3.50	1.91	37.32	-1.50	3.67	14.29	4.00	0.34	48.64	39.00
15%	-4.39	5.44	25.21	9.50	5.40	26.37	-3.94	5.17	20.89	11.00	2.14	42.19	-4.36	5.56	15.24	11.00	0.46	53.77	26.00
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-1.50	3.35	30.00	2.00	9.29	6.48	-3.11	4.45	20.97	2.00	1.78	40.11	-3.02	4.47	15.58	2.50	0.15	50.27	59.50
10%	-2.07	3.72	26.62	3.50	7.24	15.70	-2.12	3.86	21.16	5.00	2.74	36.67	-4.41	5.54	15.70	6.00	0.29	54.68	32.00
15%	-3.41	4.73	24.68	11.50	5.37	26.35	-3.43	4.75	24.46	12.00	2.06	32.24	-2.96	4.57	16.02	13.50	0.47	55.04	29.00

Table S.A.1: The table presents estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where the one is of 100,000 records and the other has 1000 records. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields. To declare a pair of records to be in the set of matches, three different thresholds are considered.

Table S.A.2 presents results across three posterior thresholds.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.29	2.22	3.30	0.40	0.12	2.28	-0.39	2.24	3.31	0.60	0.09	2.42	-0.29	2.23	3.12	0.60	0.01	2.68	60.60
10%	-0.69	2.33	2.96	1.60	0.26	4.50	-0.69	2.33	2.96	1.60	0.18	4.80	-0.60	2.31	2.79	2.40	0.11	5.80	36.00
15%	0.82	2.40	1.63	3.60	0.39	6.46	0.61	2.35	1.64	4.00	0.20	7.11	0.63	2.38	0.84	5.60	0.05	9.94	21.40
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	-0.10	2.21	3.51	1.00	0.17	1.92	0.00	2.21	3.52	1.20	0.11	2.21	0.10	2.22	3.33	1.20	0.09	2.55	57.20
10%	1.30	2.59	3.41	3.60	0.28	7.20	2.04	3.04	3.27	5.20	0.25	8.44	2.59	3.45	3.11	6.40	0.11	9.92	41.20
15%	2.75	3.59	3.18	8.60	0.76	13.53	3.10	3.87	3.21	9.60	0.53	14.39	5.18	5.69	2.93	13.80	0.13	18.37	29.00
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	-0.29	2.22	3.29	0.00	0.18	1.78	-0.29	2.22	3.29	0.00	0.18	1.78	0.10	2.22	3.33	1.20	0.02	2.88	61.20
10%	0.89	2.40	3.78	3.20	0.43	4.41	1.20	2.54	3.80	3.80	0.22	5.01	1.84	2.91	3.06	5.00	0.03	7.22	36.80
15%	1.93	2.97	3.67	5.40	0.75	11.54	2.27	3.21	3.72	6.80	0.39	12.94	2.55	3.44	3.40	9.20	0.06	16.31	27.40
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.19	2.21	3.88	0.80	0.13	2.44	-0.19	2.22	3.89	1.20	0.10	2.72	-0.39	2.25	3.91	1.60	0.08	3.05	58.20
10%	-0.89	2.40	2.78	2.20	0.31	4.80	-1.00	2.46	2.61	3.00	0.25	5.67	-1.43	2.67	2.65	4.60	0.10	7.36	41.60
15%	-1.21	2.56	4.45	5.60	0.64	11.89	-1.53	2.73	4.09	6.20	0.39	12.99	-1.06	2.54	3.60	9.00	0.09	16.34	25.00
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.39	2.23	3.86	0.40	0.16	1.61	-0.39	2.23	3.86	0.40	0.14	1.75	-0.29	2.22	3.68	0.40	0.08	2.24	56.00
10%	-0.40	2.27	3.00	3.00	0.24	5.06	-0.51	2.31	3.03	4.00	0.21	5.77	-0.10	2.27	2.45	4.60	0.09	6.94	38.60
15%	-1.12	2.52	4.48	6.20	0.36	9.32	-1.34	2.64	4.54	7.40	0.31	10.27	-1.37	2.67	4.21	9.00	0.11	12.56	24.60

Table S.A.2: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields. To declare a pair of records to be in the set of matches, three different thresholds are considered.

3

Table S.A.3 data, with three posterior threshold groups.

	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
		Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95					
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	0.19	1.78	0.37	0.25	0.01	1.32	0.19	1.78	0.37	0.25	0.00	1.41	0.19	1.78	0.37	0.25	0.00	1.41	61.38
10%	-0.13	1.78	0.63	1.62	0.18	2.59	-0.20	1.79	0.51	1.62	0.07	3.35	-0.13	1.78	0.51	1.75	0.02	4.52	38.00
15%	-0.60	1.89	0.39	3.25	0.18	6.05	-0.53	1.87	0.39	3.38	0.10	6.18	-0.36	1.84	0.39	4.75	0.02	7.63	23.25
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.06	1.77	0.75	0.88	0.22	1.63	0.12	1.78	0.63	0.88	0.08	1.74	0.50	1.85	0.51	1.50	0.01	3.12	57.38
10%	1.46	2.31	0.90	3.50	0.30	6.70	1.46	2.31	0.90	3.50	0.29	6.80	2.06	2.74	0.78	4.75	0.04	8.95	40.25
15%	3.27	3.75	0.95	8.75	0.45	10.67	3.27	3.75	0.95	8.75	0.25	11.16	4.04	4.45	0.42	10.75	0.06	14.21	30.00
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.13	1.77	1.12	0.88	0.04	1.60	0.13	1.77	1.12	0.88	0.04	1.60	0.37	1.81	0.75	1.25	0.02	3.21	60.00
10%	1.46	2.31	1.03	3.62	0.30	6.29	1.46	2.31	1.03	3.62	0.26	6.39	1.80	2.55	0.78	4.75	0.04	8.26	39.75
15%	2.73	3.29	0.68	8.50	0.54	11.89	2.73	3.29	0.68	8.50	0.45	12.27	4.23	4.63	0.42	12.00	0.04	16.44	29.25
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.06	1.77	0.75	0.38	0.24	1.22	-0.06	1.77	0.75	0.38	0.04	1.33	-0.07	1.77	0.50	0.88	0.00	3.03	60.12
10%	-0.00	1.77	1.01	1.50	0.35	3.23	-0.00	1.77	1.01	1.50	0.21	3.81	-0.08	1.79	0.51	2.62	0.05	5.95	38.75
15%	-1.02	2.08	0.92	6.00	0.58	10.44	-0.96	2.05	0.79	6.00	0.41	10.62	-1.28	2.25	0.54	8.75	0.08	14.56	26.12
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.06	1.77	0.87	0.50	0.18	0.94	-0.19	1.78	0.75	0.62	0.03	1.61	-0.13	1.78	0.50	0.75	0.01	2.70	60.38
10%	-0.27	1.80	0.76	2.50	0.23	4.65	-0.20	1.80	0.64	2.50	0.13	4.87	-0.41	1.84	0.52	3.75	0.05	6.60	40.38
15%	-1.04	2.10	1.07	7.38	0.39	9.71	-1.11	2.14	1.07	7.50	0.32	10.02	-0.89	2.06	0.55	9.75	0.04	12.96	23.75

Table S.A.3: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields. To declare a pair of records to be in the set of matches, three different thresholds are considered.

4

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.87	1.34	16.74	0.80	2.35	28.12	-0.84	1.32	16.55	0.95	1.99	28.75	-1.01	1.46	12.20	1.05	0.15	39.08	58.15
10%	-1.18	1.57	17.06	2.50	3.50	27.04	-1.07	1.51	12.31	2.75	0.53	38.54	-1.03	1.49	10.31	3.40	0.17	41.31	35.25
15%	-0.59	1.23	11.59	4.65	1.32	35.43	-0.76	1.33	10.98	5.55	1.11	36.25	-0.79	1.36	8.78	7.05	0.26	41.19	21.75
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	-0.68	1.19	24.79	0.80	1.88	10.92	-1.03	1.45	17.58	0.85	1.84	30.89	-0.51	1.17	12.70	1.05	0.16	41.06	58.75
10%	0.45	1.10	23.43	4.90	2.42	20.62	0.82	1.32	18.02	5.40	2.24	33.85	1.09	1.53	12.34	6.20	0.20	44.89	42.20
15%	1.81	2.09	18.92	9.55	3.48	34.30	2.81	3.02	14.33	11.80	0.97	43.35	3.29	3.47	11.68	13.00	0.25	48.54	29.55
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	-1.29	1.62	25.06	0.70	1.99	11.20	-0.89	1.36	17.18	0.95	1.87	32.03	-0.38	1.12	12.92	1.30	0.16	42.96	59.65
10%	-0.26	1.03	23.81	5.10	6.49	18.62	-0.08	1.04	18.50	5.50	2.41	33.57	0.39	1.15	12.09	6.60	0.17	48.05	40.70
15%	2.18	2.43	18.47	9.75	3.91	32.71	3.12	3.31	13.93	11.65	1.06	40.88	3.12	3.31	11.17	12.95	0.31	45.75	29.90
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-2.02	2.24	24.32	0.90	1.93	10.50	-1.41	1.74	16.81	1.05	1.87	30.26	-1.43	1.78	12.56	1.15	0.15	41.40	57.60
10%	-1.37	1.72	18.54	4.45	2.89	30.99	-1.42	1.76	17.99	5.15	2.68	31.74	-1.87	2.16	11.18	6.25	0.19	45.18	38.85
15%	-1.98	2.25	18.27	8.05	3.93	31.77	-2.65	2.87	13.55	10.35	1.07	41.89	-3.00	3.21	10.72	11.70	0.28	47.59	26.30
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-1.85	2.09	24.37	0.85	1.74	10.35	-1.29	1.64	16.94	0.95	1.28	32.04	-1.66	1.96	13.05	1.10	0.16	40.67	57.70
10%	-1.45	1.78	17.63	3.55	3.03	29.83	-1.64	1.95	16.30	4.50	1.59	34.35	-2.02	2.29	11.93	5.10	0.21	43.85	39.05
15%	-2.36	2.59	18.60	8.55	3.91	31.53	-3.08	3.27	14.29	11.20	1.01	41.12	-3.49	3.67	10.88	12.35	0.25	46.23	26.60

Table S.A.4: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is considered. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.12	0.71	2.35	0.32	0.20	1.75	-0.10	0.71	2.20	0.42	0.16	2.18	-0.10	0.71	2.05	0.54	0.02	2.57	59.70
10%	-0.08	0.71	2.09	1.56	0.27	3.98	0.02	0.71	1.82	1.86	0.19	4.56	0.15	0.73	0.87	2.38	0.03	6.87	35.32
15%	-0.31	0.78	1.89	3.46	0.52	7.86	-0.34	0.79	1.74	4.26	0.28	8.95	-0.31	0.79	0.90	5.68	0.06	11.42	20.90
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.26	0.75	2.38	0.62	0.25	2.51	0.32	0.77	2.21	0.78	0.14	2.99	0.45	0.84	2.16	1.10	0.10	3.38	59.66
10%	1.09	1.30	2.70	3.30	0.47	7.19	1.37	1.55	2.42	4.02	0.22	8.31	1.77	1.91	2.29	5.40	0.10	9.68	40.92
15%	3.18	3.26	2.77	8.02	0.80	14.66	3.53	3.60	2.53	9.22	0.42	16.10	4.52	4.58	2.09	12.12	0.09	19.46	29.26
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.10	0.71	2.43	0.30	0.28	2.32	0.28	0.75	2.28	0.60	0.16	2.90	0.45	0.84	2.16	1.10	0.10	3.50	58.82
10%	1.34	1.51	2.51	3.56	0.47	7.11	1.62	1.77	2.25	4.32	0.21	8.32	2.02	2.14	2.18	5.58	0.09	9.60	39.14
15%	2.57	2.67	3.07	8.42	0.87	12.94	2.93	3.02	2.80	9.72	0.38	14.50	3.84	3.91	2.28	12.44	0.09	17.94	28.56
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.09	0.71	2.41	0.36	0.24	2.21	-0.08	0.71	2.30	0.48	0.15	2.59	-0.13	0.71	2.19	0.82	0.11	3.12	56.86
10%	-0.27	0.76	2.32	2.30	0.47	6.44	-0.26	0.75	2.16	2.92	0.24	7.44	-0.58	0.92	1.95	4.46	0.09	9.10	37.50
15%	-0.70	1.01	2.16	6.62	0.75	13.12	-0.90	1.16	1.81	7.96	0.34	14.86	-1.22	1.43	1.70	10.80	0.09	17.66	26.32
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.03	0.70	2.35	0.26	0.24	2.45	-0.03	0.70	2.20	0.38	0.14	2.90	-0.18	0.72	2.15	0.76	0.10	3.35	58.20
10%	-0.67	0.98	2.49	3.04	0.47	6.59	-0.74	1.03	2.21	3.74	0.22	7.75	-1.03	1.26	2.10	5.04	0.10	9.11	38.24
15%	-1.23	1.42	2.60	6.96	0.85	13.64	-1.58	1.74	2.33	8.42	0.38	15.28	-2.23	2.35	1.90	11.24	0.08	18.70	26.66

Table S.A.5: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate $\widehat{FDR}$, the estimated false discovery rate $\widehat{FNR}$, and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	0.02	0.56	0.46	0.31	0.05	1.06	0.04	0.56	0.23	0.33	0.05	2.24	0.06	0.56	0.19	0.56	0.01	2.64	59.25
10%	0.03	0.56	0.46	1.58	0.27	3.06	0.02	0.56	0.43	1.61	0.11	3.31	-0.06	0.57	0.27	2.28	0.02	4.92	36.10
15%	-0.01	0.57	0.45	3.80	0.23	6.80	0.00	0.57	0.41	3.80	0.20	6.89	0.03	0.57	0.30	5.15	0.03	8.45	22.43
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.31	0.64	0.54	0.67	0.24	1.55	0.32	0.65	0.54	0.70	0.22	1.64	0.52	0.76	0.24	1.14	0.01	3.69	59.71
10%	1.25	1.37	0.67	3.25	0.38	4.74	1.25	1.37	0.63	3.24	0.33	4.95	1.77	1.86	0.29	4.62	0.02	7.62	40.60
15%	3.52	3.57	0.83	8.47	0.51	11.89	3.59	3.63	0.68	8.46	0.43	12.23	4.46	4.50	0.42	11.05	0.07	15.55	29.34
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.11	0.57	0.67	0.51	0.25	1.41	0.15	0.58	0.60	0.51	0.19	1.87	0.35	0.66	0.24	0.99	0.01	4.09	60.39
10%	1.24	1.36	0.76	3.71	0.37	4.98	1.23	1.35	0.68	3.74	0.32	5.31	1.74	1.83	0.37	5.17	0.05	8.03	40.10
15%	2.98	3.04	0.76	8.21	0.53	10.35	3.01	3.06	0.56	8.20	0.40	10.88	3.75	3.80	0.33	10.53	0.08	14.04	28.09
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.07	0.56	0.60	0.44	0.24	1.22	-0.07	0.56	0.60	0.47	0.22	1.30	-0.04	0.56	0.19	0.89	0.01	3.42	59.04
10%	-0.16	0.58	0.57	2.28	0.35	4.03	-0.16	0.59	0.53	2.28	0.31	4.17	-0.29	0.64	0.23	3.54	0.02	6.58	37.52
15%	-0.59	0.83	0.56	6.24	0.49	9.52	-0.59	0.83	0.56	6.26	0.46	9.61	-1.12	1.26	0.22	8.69	0.09	12.93	26.29
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.04	0.56	0.57	0.41	0.27	1.16	-0.03	0.56	0.52	0.46	0.17	1.42	-0.05	0.56	0.28	0.85	0.01	3.36	57.96
10%	-0.34	0.66	0.74	2.47	0.38	4.53	-0.35	0.66	0.70	2.49	0.29	4.72	-0.65	0.86	0.34	4.10	0.02	7.87	38.17
15%	-1.00	1.16	0.72	6.75	0.57	9.75	-0.99	1.14	0.59	6.75	0.46	10.22	-1.38	1.50	0.23	9.40	0.09	13.46	26.79

Table S.A.6: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior $\geq$ 0.75						Threshold 2: Posterior $\geq$ 0.85						Threshold 3: Posterior $\geq$ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.39	0.51	12.85	0.73	2.49	26.33	-0.29	0.45	9.87	0.88	0.35	34.86	-0.26	0.43	9.19	1.09	0.17	36.32	54.92
10%	-0.39	0.52	13.14	2.62	3.33	24.05	-0.24	0.42	9.45	3.62	0.74	34.12	-0.25	0.43	7.83	4.31	0.21	37.64	32.09
15%	-0.32	0.47	13.58	6.50	4.93	28.09	-0.31	0.47	9.82	8.69	1.16	39.00	-0.20	0.41	7.25	10.00	0.24	44.72	17.94
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	-0.19	0.38	12.94	1.04	2.25	27.25	0.06	0.34	10.29	1.31	0.35	34.81	0.23	0.41	9.69	1.56	0.19	36.17	58.01
10%	1.00	1.05	13.76	4.61	3.29	28.24	1.60	1.63	10.46	5.83	0.72	36.74	2.03	2.06	8.97	6.66	0.22	40.35	39.30
15%	3.33	3.35	12.94	9.72	2.81	34.15	4.09	4.11	10.37	12.23	1.09	39.78	4.77	4.79	8.16	13.66	0.29	44.35	27.84
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	-0.17	0.37	13.18	1.36	2.28	27.32	0.08	0.34	11.40	1.56	0.38	33.27	0.29	0.44	9.79	1.84	0.18	36.64	57.63
10%	0.95	1.01	14.07	4.61	3.50	27.00	1.68	1.71	10.59	5.88	0.74	36.20	1.97	2.00	9.03	6.61	0.22	39.98	39.24
15%	3.09	3.11	13.85	10.09	3.32	31.94	3.66	3.68	10.63	12.31	1.05	39.49	4.22	4.24	8.65	13.57	0.28	44.25	27.92
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.45	0.56	13.10	1.03	1.92	29.54	-0.48	0.59	12.13	1.17	0.88	32.62	-0.61	0.70	9.61	1.39	0.18	37.91	56.89
10%	-0.79	0.85	13.88	3.67	3.25	28.35	-1.01	1.07	10.13	4.74	0.62	37.45	-1.17	1.22	8.66	5.45	0.22	40.52	36.66
15%	-1.44	1.48	14.22	8.64	3.97	33.00	-1.80	1.83	10.74	11.44	1.01	41.46	-2.13	2.16	8.44	13.00	0.27	46.16	25.15
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.56	0.65	13.05	1.10	2.37	26.92	-0.68	0.76	11.44	1.27	1.02	32.47	-0.83	0.89	9.78	1.50	0.18	36.68	56.84
10%	-0.92	0.98	13.91	4.03	3.37	27.24	-1.18	1.23	10.37	5.24	0.67	36.53	-1.50	1.54	8.82	5.97	0.22	39.88	37.51
15%	-1.91	1.94	13.16	9.24	2.68	32.76	-2.20	2.23	10.54	11.59	1.10	38.24	-2.66	2.68	8.26	12.81	0.29	43.12	26.29

Table S.A.7: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets with of 100,000 records each. In addition, there is a 20% overlap and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.03	0.22	1.09	0.52	0.24	2.17	-0.02	0.22	1.02	0.63	0.15	2.60	-0.02	0.22	0.94	0.92	0.11	3.18	54.95
10%	-0.00	0.22	1.12	1.83	0.45	4.51	0.01	0.23	0.98	2.34	0.22	5.49	-0.04	0.23	0.89	3.55	0.10	6.92	31.72
15%	-0.09	0.24	1.18	4.97	0.72	9.33	-0.05	0.23	0.92	5.99	0.33	11.06	-0.03	0.23	0.53	8.51	0.05	14.44	17.84
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.24	0.32	1.26	0.74	0.23	3.01	0.33	0.40	1.20	0.93	0.13	3.50	0.46	0.52	1.18	1.37	0.10	3.99	58.17
10%	1.39	1.41	1.35	3.48	0.53	7.32	1.67	1.68	1.25	4.15	0.25	8.27	2.20	2.22	1.06	5.69	0.09	10.24	39.78
15%	3.45	3.46	1.51	8.80	0.75	14.22	3.86	3.87	1.30	9.92	0.38	15.60	4.65	4.66	0.99	12.31	0.10	18.58	28.01
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.21	0.31	1.24	0.83	0.21	3.06	0.28	0.36	1.19	1.00	0.13	3.44	0.43	0.48	1.16	1.42	0.09	3.96	58.34
10%	1.29	1.31	1.34	3.53	0.48	6.68	1.55	1.57	1.26	4.28	0.28	7.57	2.09	2.11	1.08	5.74	0.09	9.54	39.67
15%	3.18	3.19	1.54	8.71	0.74	13.06	3.49	3.50	1.38	9.76	0.44	14.33	4.14	4.15	1.06	12.07	0.10	17.34	28.25
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.10	0.25	1.24	0.58	0.20	2.83	-0.13	0.26	1.14	0.70	0.13	3.21	-0.19	0.29	1.12	1.09	0.09	3.67	56.35
10%	-0.31	0.38	1.26	2.68	0.44	6.50	-0.38	0.44	1.09	3.30	0.21	7.55	-0.57	0.61	1.01	4.75	0.09	9.04	37.01
15%	-0.68	0.72	1.36	6.88	0.78	12.66	-0.87	0.90	1.20	8.11	0.42	14.09	-1.16	1.19	0.94	10.81	0.10	17.36	25.37
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.12	0.25	1.21	0.59	0.21	2.60	-0.14	0.26	1.13	0.71	0.13	2.94	-0.20	0.30	1.11	1.11	0.10	3.39	56.99
10%	-0.48	0.53	1.34	2.79	0.46	6.10	-0.59	0.63	1.18	3.46	0.22	7.24	-0.80	0.83	1.07	5.00	0.09	8.99	37.78
15%	-1.22	1.24	1.38	7.24	0.75	12.19	-1.42	1.44	1.23	8.47	0.44	13.50	-1.94	1.95	1.01	11.24	0.10	16.73	26.31

Table S.A.8: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets with of 100,000 records each. In addition, there is a 50% overlap and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields. To declare a pair of records to be in the set of matches, three different thresholds are considered.

Table S.A.9

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.01	0.18	0.14	0.43	0.21	1.37	-0.01	0.18	0.13	0.47	0.19	1.57	0.00	0.18	0.05	0.69	0.02	3.28	54.57
10%	-0.00	0.18	0.15	1.83	0.29	3.66	0.00	0.18	0.14	1.86	0.26	3.79	0.01	0.18	0.08	2.87	0.05	5.65	31.56
15%	0.00	0.18	0.23	4.83	0.49	7.62	0.01	0.18	0.20	4.91	0.39	8.17	-0.01	0.18	0.12	7.28	0.09	11.06	17.91
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.27	0.32	0.14	0.70	0.22	1.81	0.27	0.32	0.13	0.75	0.18	2.01	0.41	0.44	0.05	1.11	0.02	4.15	58.02
10%	1.39	1.40	0.19	3.57	0.38	5.78	1.40	1.41	0.17	3.62	0.29	6.17	1.96	1.97	0.08	5.06	0.05	8.85	39.32
15%	3.57	3.57	0.24	8.53	0.50	12.56	3.58	3.59	0.20	8.56	0.41	12.96	4.40	4.40	0.10	11.04	0.08	16.34	27.78
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.25	0.30	0.14	0.69	0.22	1.80	0.24	0.30	0.13	0.73	0.18	2.02	0.39	0.43	0.04	1.10	0.02	4.20	58.07
10%	1.33	1.34	0.20	3.62	0.34	5.48	1.32	1.33	0.18	3.64	0.28	5.80	1.75	1.76	0.09	5.02	0.05	8.47	39.55
15%	3.27	3.27	0.24	8.73	0.50	11.48	3.27	3.28	0.23	8.79	0.38	11.80	3.86	3.86	0.10	11.11	0.09	15.30	28.05
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.05	0.18	0.13	0.49	0.20	1.66	-0.05	0.18	0.13	0.54	0.18	1.88	-0.08	0.19	0.05	0.83	0.02	3.86	56.46
10%	-0.32	0.37	0.17	2.53	0.34	5.08	-0.32	0.37	0.15	2.56	0.28	5.37	-0.52	0.55	0.08	3.89	0.05	7.89	36.79
15%	-0.76	0.78	0.24	6.64	0.55	10.96	-0.77	0.79	0.23	6.70	0.43	11.23	-1.11	1.12	0.12	9.47	0.09	15.00	25.41
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.07	0.19	0.16	0.59	0.22	1.68	-0.06	0.19	0.14	0.63	0.17	1.91	-0.09	0.20	0.06	0.94	0.02	3.99	57.00
10%	-0.38	0.42	0.19	2.77	0.36	4.78	-0.38	0.42	0.15	2.78	0.27	5.18	-0.61	0.63	0.07	4.18	0.05	7.82	37.53
15%	-1.24	1.25	0.28	7.37	0.54	10.48	-1.24	1.25	0.26	7.43	0.42	10.76	-1.69	1.70	0.12	10.04	0.09	14.45	26.32

Table S.A.9: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets with of 100,000 records each. In addition, there is a 80% overlap and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-2.25	3.81	25.94	1.50	0.27	10.87	-3.06	4.53	12.44	1.50	0.21	49.17	-3.99	5.22	12.67	3.50	0.03	51.57	57.50
10%	-0.24	3.26	21.43	6.50	7.05	15.43	-2.92	4.51	12.74	7.50	0.17	45.58	-3.16	4.68	12.80	8.00	0.15	45.81	40.50
15%	-0.62	3.38	20.18	9.00	1.08	27.54	-2.26	4.23	7.61	9.00	0.70	45.05	-2.31	4.35	6.49	13.50	0.18	48.06	28.00
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.37	3.08	27.61	3.00	0.20	10.01	-1.28	3.56	14.54	3.00	0.17	44.42	-1.06	3.50	14.60	3.50	0.13	44.73	60.50
10%	2.07	3.77	25.79	6.50	9.68	7.67	-0.35	3.42	13.82	6.50	0.28	46.61	-0.56	3.49	14.15	9.00	0.17	47.80	47.50
15%	2.58	4.16	24.89	12.50	11.27	8.76	2.80	4.61	10.75	17.00	0.21	50.57	3.45	5.05	10.99	19.00	0.16	51.39	28.50
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	-0.02	3.05	28.15	3.00	0.38	9.48	-0.84	3.42	15.28	3.00	0.34	47.89	0.54	3.41	15.84	7.00	0.18	49.48	59.50
10%	1.26	3.38	25.59	5.50	9.86	11.65	0.84	3.52	12.15	6.00	0.26	49.43	1.08	3.60	12.21	6.50	0.24	49.63	43.00
15%	1.06	3.44	22.65	9.50	8.42	11.62	1.83	4.02	12.31	14.50	0.27	48.40	1.59	3.93	11.86	14.50	0.23	48.86	30.00
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-1.31	3.33	26.59	2.00	0.22	7.06	-1.50	3.63	14.78	2.00	0.19	47.43	-1.28	3.56	14.98	3.50	0.02	49.56	62.50
10%	-2.28	3.87	26.56	6.00	8.36	14.10	-3.39	4.83	11.79	6.50	0.24	47.71	-3.63	5.01	11.85	7.00	0.22	47.93	39.00
15%	-5.12	6.07	25.96	13.00	9.87	11.72	-6.64	7.58	12.43	19.00	0.17	49.39	-7.37	8.25	12.85	22.00	0.15	50.23	26.00
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-2.05	3.65	29.82	3.50	0.24	8.04	-2.60	4.22	14.98	3.50	0.18	48.07	-3.06	4.53	15.11	4.50	0.03	50.04	59.50
10%	-3.88	5.00	26.98	8.00	0.34	15.41	-4.77	5.87	14.02	8.00	0.22	49.89	-5.02	6.08	14.08	8.50	0.21	50.10	32.00
15%	-4.13	5.30	25.00	14.50	9.67	11.26	-6.64	7.58	12.97	19.50	0.16	50.06	-7.54	8.40	13.19	21.00	0.15	50.55	29.00

Table S.A.10: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets where the one is of 100,000 records and the other has 1000 records. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields To declare a pair of records to be in the set of matches, three different thresholds are considered. For each linkage field, missing comparisons have been coded as disagreements.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.29	2.22	3.50	0.60	0.16	2.81	-0.29	2.22	3.50	0.60	0.16	2.81	0.39	2.26	3.35	1.80	0.05	4.50	60.60
10%	-0.30	2.26	3.59	3.40	0.40	10.87	-0.30	2.27	3.03	4.00	0.15	12.56	-0.41	2.29	3.05	4.60	0.14	13.01	36.00
15%	1.23	2.58	2.88	5.60	0.31	18.46	1.23	2.58	2.88	5.60	0.31	18.46	1.28	2.64	1.91	7.80	0.14	20.95	21.40
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.39	2.26	3.35	1.80	0.10	3.55	0.59	2.30	3.36	2.20	0.07	3.85	0.79	2.37	3.37	2.60	0.05	4.19	57.20
10%	2.60	3.46	4.16	7.80	0.36	12.75	3.04	3.80	3.98	8.40	0.30	13.30	5.07	5.58	3.08	12.00	0.05	17.19	41.20
15%	5.36	5.85	4.46	14.40	0.83	23.58	6.55	6.97	3.91	16.40	0.38	25.76	7.24	7.63	3.50	17.40	0.14	27.41	29.00
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.29	2.23	3.70	1.20	0.10	3.93	0.30	2.24	3.55	2.20	0.09	4.82	0.30	2.24	3.55	2.20	0.09	4.82	61.20
10%	1.74	2.85	4.91	7.00	0.34	10.23	1.86	2.94	4.75	7.80	0.19	10.95	2.42	3.34	3.79	8.60	0.09	12.88	36.80
15%	3.43	4.14	4.51	11.00	0.51	30.29	3.46	4.17	4.33	11.60	0.31	31.32	3.74	4.42	4.19	13.00	0.13	32.93	27.40
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.67	2.29	5.76	1.80	0.27	3.60	-0.20	2.22	4.31	2.40	0.07	5.70	-0.20	2.22	4.31	2.40	0.07	5.70	58.20
10%	-1.41	2.65	4.02	4.40	0.62	22.20	-1.53	2.73	3.48	5.60	0.11	24.08	-2.11	3.12	3.59	8.60	0.10	25.76	41.60
15%	-1.28	2.64	4.89	10.60	0.38	23.87	-1.50	2.75	4.70	10.80	0.32	24.41	-1.73	2.90	4.55	11.80	0.19	25.68	25.00
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.39	2.23	3.86	0.40	0.12	4.11	-0.39	2.23	3.86	0.40	0.12	4.11	-0.39	2.26	3.94	2.40	0.05	5.73	56.00
10%	-1.13	2.53	3.70	6.20	0.19	12.30	-1.13	2.53	3.70	6.20	0.19	12.30	-1.16	2.57	2.95	7.80	0.07	14.84	38.60
15%	-2.32	3.23	9.29	10.20	0.75	20.18	-1.58	2.79	5.47	10.20	0.65	33.08	-2.32	3.30	4.86	13.80	0.11	36.56	24.60

Table S.A.11: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields To declare a pair of records to be in the set of matches, three different thresholds are considered. For each linkage field, missing comparisons have been coded as disagreements.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	0.13	1.77	0.75	0.50	0.16	2.20	0.13	1.77	0.75	0.50	0.16	2.20	-0.13	1.78	0.75	1.25	0.03	3.78	61.38
10%	-0.07	1.78	0.76	2.12	0.56	7.51	-0.02	1.79	0.51	3.25	0.13	9.99	-0.02	1.79	0.51	3.25	0.13	9.99	38.00
15%	-0.03	1.81	0.65	5.12	0.50	11.62	-0.30	1.84	0.40	5.88	0.19	13.25	-0.26	1.86	0.41	8.25	0.07	15.29	23.25
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.75	1.93	0.63	2.12	0.15	2.88	0.75	1.93	0.63	2.12	0.15	2.88	1.20	2.16	0.64	3.25	0.05	4.11	57.38
10%	3.13	3.62	1.19	6.38	0.42	6.88	3.28	3.76	0.93	6.88	0.12	8.75	3.28	3.76	0.93	6.88	0.12	8.75	40.25
15%	4.59	4.96	1.27	12.88	0.68	12.26	4.54	4.91	0.85	12.88	0.33	13.00	5.79	6.09	0.88	16.00	0.07	17.00	30.00
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.81	1.95	1.13	2.00	0.17	2.76	0.94	2.01	1.14	2.25	0.14	2.96	1.33	2.22	1.15	3.00	0.05	3.87	60.00
10%	1.95	2.67	0.93	6.38	0.31	8.67	2.29	2.93	0.93	7.00	0.14	9.30	2.29	2.93	0.93	7.00	0.11	10.62	39.75
15%	5.00	5.35	1.17	15.25	0.44	13.24	5.26	5.60	0.88	15.88	0.18	14.52	5.63	5.95	0.30	16.62	0.08	15.93	29.25
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.13	1.77	0.75	0.75	0.19	1.92	-0.06	1.77	0.75	0.88	0.17	2.12	-0.33	1.82	0.76	2.62	0.04	3.95	60.12
10%	-0.21	1.81	0.90	4.00	0.38	6.97	-0.35	1.84	0.78	4.62	0.10	8.82	-0.35	1.84	0.78	4.62	0.10	8.82	38.75
15%	-1.74	2.55	0.97	10.88	0.45	22.89	-1.74	2.55	0.97	10.88	0.45	22.89	-2.27	2.96	0.72	14.12	0.10	26.41	26.12
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	0.06	1.77	0.75	0.88	0.20	2.44	0.06	1.77	0.75	1.12	0.09	2.76	0.12	1.79	0.76	2.50	0.05	4.08	60.38
10%	-0.83	2.00	0.53	5.75	0.10	6.10	-0.83	2.00	0.53	5.75	0.10	6.10	-0.83	2.00	0.53	5.75	0.10	6.10	40.38
15%	-0.62	1.96	1.38	10.75	0.53	15.51	-1.43	2.37	0.71	12.50	0.25	17.66	-2.76	3.36	0.74	15.62	0.08	20.61	23.75

Table S.A.12: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields To declare a pair of records to be in the set of matches, three different thresholds are considered. For each linkage field, missing comparisons have been coded as disagreements.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-1.14	1.50	24.93	1.70	0.24	14.84	-0.85	1.35	12.51	1.70	0.22	41.40	-1.01	1.46	12.17	2.55	0.13	42.18	58.15
10%	-1.35	1.68	22.89	3.80	7.79	17.34	-0.56	1.22	10.97	6.25	0.23	45.00	-0.58	1.23	10.76	6.25	0.21	45.21	35.25
15%	-0.58	1.18	20.35	6.85	1.33	28.97	-0.45	1.21	9.48	9.30	0.58	46.18	-0.67	1.32	8.77	12.60	0.21	48.66	21.75
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.02	0.98	24.97	2.35	0.32	14.02	0.04	1.05	13.24	2.35	0.29	41.51	0.36	1.12	13.00	4.00	0.15	42.70	58.75
10%	1.40	1.73	23.99	8.60	7.72	14.50	1.88	2.18	12.44	9.20	0.27	43.59	2.17	2.43	12.46	9.70	0.25	43.91	42.20
15%	2.69	2.89	22.92	12.75	8.05	16.76	4.07	4.22	11.62	14.85	0.39	43.37	4.25	4.40	11.27	17.35	0.22	44.91	29.55
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	-0.58	1.14	25.12	2.50	0.28	14.22	0.09	1.06	13.06	2.50	0.24	43.12	0.10	1.07	12.77	3.35	0.15	43.90	59.65
10%	0.15	1.02	24.45	7.00	8.30	14.23	1.05	1.52	12.05	9.10	0.24	45.48	1.41	1.79	12.05	10.20	0.21	46.01	40.70
15%	2.92	3.10	22.72	12.60	1.46	23.88	3.83	3.99	11.12	15.25	0.60	44.04	4.33	4.48	10.88	17.65	0.22	45.96	29.90
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-2.07	2.29	24.31	1.90	0.26	14.40	-1.47	1.81	12.68	1.90	0.24	42.00	-1.68	1.99	12.51	3.10	0.14	42.90	57.60
10%	-2.38	2.59	23.78	8.50	0.57	21.16	-2.09	2.36	11.50	9.15	0.27	43.79	-2.18	2.45	11.49	9.85	0.25	44.18	38.85
15%	-3.31	3.48	21.68	11.65	1.43	22.85	-3.49	3.67	10.59	14.35	0.60	43.23	-3.81	3.98	10.59	16.80	0.26	45.08	26.30
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-1.94	2.17	24.97	2.20	0.27	13.89	-1.72	2.02	13.45	2.20	0.24	41.58	-1.89	2.17	13.28	3.05	0.16	42.30	57.70
10%	-2.65	2.84	23.09	7.25	0.65	20.81	-2.29	2.54	12.00	8.00	0.28	43.44	-2.51	2.74	11.95	8.60	0.25	43.83	39.05
15%	-4.15	4.28	22.56	12.10	1.28	24.51	-3.57	3.75	10.78	13.10	0.58	43.06	-4.05	4.21	10.58	16.75	0.22	45.41	26.60

Table S.A.13: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields To declare a pair of records to be in the set of matches, three different thresholds are considered. For each linkage field, missing comparisons have been coded as disagreements.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.20	0.73	2.78	0.68	0.31	3.75	-0.12	0.71	2.73	0.92	0.17	4.01	-0.09	0.71	2.21	1.90	0.09	6.20	59.70
10%	-0.01	0.71	2.93	3.14	0.51	9.49	-0.00	0.71	2.31	3.70	0.16	11.49	0.06	0.72	2.24	3.90	0.15	11.75	35.32
15%	-0.42	0.84	3.07	6.68	0.63	22.24	-0.41	0.83	2.99	6.70	0.39	22.68	-0.54	0.91	2.17	9.82	0.11	26.45	20.90
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.54	0.89	2.84	1.44	0.33	4.17	0.63	0.94	2.81	1.64	0.20	4.39	0.99	1.22	2.82	2.84	0.14	5.41	59.66
10%	2.11	2.22	3.53	6.50	0.46	9.47	2.25	2.36	3.39	7.08	0.22	10.18	2.55	2.66	2.72	7.56	0.10	12.25	40.92
15%	4.93	4.99	3.37	13.38	0.64	20.72	5.37	5.42	3.19	14.32	0.41	21.68	5.79	5.84	2.99	15.64	0.16	23.27	29.26
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.48	0.85	2.92	1.66	0.33	4.73	0.58	0.91	2.83	1.84	0.30	4.97	0.97	1.20	2.79	3.18	0.10	6.25	58.82
10%	2.25	2.36	3.50	6.68	0.42	11.43	2.47	2.57	3.44	7.36	0.20	12.19	2.55	2.65	2.71	7.36	0.17	13.74	39.14
15%	4.09	4.15	3.54	13.40	0.66	21.97	4.41	4.48	3.33	14.60	0.38	23.20	4.61	4.67	3.19	15.70	0.17	24.63	28.56
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.20	0.73	2.89	1.04	0.26	4.95	-0.14	0.71	2.83	1.14	0.25	5.07	-0.25	0.75	2.84	2.22	0.08	6.09	56.86
10%	-0.83	1.10	3.38	5.70	0.55	14.63	-0.89	1.15	3.21	6.44	0.21	15.76	-0.80	1.08	2.56	6.44	0.11	17.50	37.50
15%	-1.51	1.68	3.47	11.98	0.61	24.26	-1.50	1.67	3.36	11.98	0.37	24.62	-2.04	2.17	2.93	14.48	0.16	27.11	26.32
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.34	0.78	2.86	0.98	0.28	4.97	-0.38	0.80	2.81	1.08	0.16	5.14	-0.48	0.85	2.78	2.18	0.11	6.16	58.20
10%	-1.37	1.55	3.52	6.20	0.61	13.17	-1.55	1.71	3.28	6.90	0.19	14.30	-1.44	1.61	2.49	6.90	0.17	16.05	38.24
15%	-2.47	2.58	3.44	12.32	0.57	26.25	-2.50	2.61	3.29	12.32	0.36	26.65	-3.04	3.13	3.14	14.98	0.16	28.99	26.66

Table S.A.14: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate $\widehat{\text{FDR}}$, the estimated false discovery rate $\widehat{\text{FNR}}$, and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields To declare a pair of records to be in the set of matches, three different thresholds are considered. For each linkage field, missing comparisons have been coded as disagreements.

	Threshold 1: Posterior $\geq$ 0.75						Threshold 2: Posterior $\geq$ 0.85						Threshold 3: Posterior $\geq$ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	0.07	0.56	0.58	0.75	0.21	2.02	0.06	0.56	0.58	0.83	0.17	2.12	0.10	0.57	0.46	1.96	0.04	3.93	59.25
10%	-0.04	0.57	0.54	2.67	0.25	6.54	-0.07	0.57	0.49	3.70	0.12	7.70	-0.07	0.57	0.49	3.70	0.11	7.73	36.10
15%	-0.07	0.58	0.88	5.53	0.49	11.95	0.01	0.58	0.60	6.71	0.18	13.92	0.07	0.59	0.55	9.20	0.09	16.15	22.43
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.70	0.90	0.69	1.62	0.25	2.28	0.76	0.95	0.70	1.88	0.17	2.54	1.15	1.28	0.65	2.94	0.06	3.72	59.71
10%	2.57	2.63	0.84	6.49	0.42	5.48	2.80	2.86	0.75	7.21	0.16	7.02	2.80	2.86	0.75	7.21	0.13	7.07	40.60
15%	5.00	5.03	0.78	12.90	0.34	10.65	5.37	5.40	0.75	14.06	0.19	11.85	5.75	5.78	0.66	15.20	0.10	13.44	29.34
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.50	0.75	0.78	1.50	0.24	2.40	0.52	0.77	0.77	1.60	0.18	2.55	0.87	1.03	0.74	2.72	0.05	3.76	60.39
10%	2.55	2.61	0.85	7.08	0.39	6.19	2.77	2.83	0.73	7.64	0.15	7.69	2.77	2.83	0.73	7.64	0.12	7.74	40.10
15%	4.28	4.32	0.85	12.74	0.51	13.20	4.27	4.31	0.84	12.74	0.31	13.41	4.75	4.79	0.60	14.95	0.10	16.30	28.09
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.16	0.58	0.72	1.17	0.19	2.62	-0.16	0.58	0.72	1.33	0.18	2.76	-0.25	0.62	0.69	2.38	0.05	3.89	59.04
10%	-0.64	0.86	0.79	5.26	0.39	7.08	-0.66	0.87	0.62	5.72	0.15	8.51	-0.66	0.87	0.62	5.72	0.15	8.52	37.52
15%	-1.61	1.72	0.89	11.21	0.47	15.01	-1.60	1.70	0.81	11.21	0.29	15.34	-2.01	2.09	0.59	13.56	0.10	18.31	26.29
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.21	0.60	0.69	1.40	0.20	2.75	-0.23	0.61	0.68	1.49	0.16	2.87	-0.30	0.64	0.65	2.62	0.05	4.08	57.96
10%	-0.95	1.11	0.91	5.90	0.35	7.72	-1.10	1.24	0.93	6.49	0.13	8.32	-1.08	1.23	0.74	6.49	0.13	9.36	38.17
15%	-1.85	1.95	0.86	11.70	0.50	15.48	-1.85	1.95	0.80	11.70	0.30	15.89	-2.27	2.34	0.58	14.31	0.10	19.05	26.79

Table S.A.15: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields To declare a pair of records to be in the set of matches, three different thresholds are considered. For each linkage field, missing comparisons have been coded as disagreements.

Table S.A.16: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets with of 100,000 records each. In addition, there is a 20% overlap and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields To declare a pair of records to be in the set of matches, three different thresholds are considered. For each linkage field, missing comparisons have been coded as disagreements.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.46	0.56	19.75	2.08	8.21	11.06	-0.20	0.40	9.53	2.08	0.23	39.88	-0.19	0.39	9.43	3.22	0.14	40.82	54.92
10%	-0.41	0.52	18.75	4.03	8.44	16.93	-0.29	0.45	8.11	7.03	0.22	43.53	-0.29	0.45	8.11	7.03	0.22	43.53	32.09
15%	-0.13	0.38	8.53	9.72	0.85	43.21	-0.26	0.44	8.01	10.78	0.77	43.89	-0.21	0.43	7.20	14.95	0.22	47.23	17.94
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.53	0.62	19.83	2.75	0.28	14.59	0.60	0.69	10.04	2.75	0.25	38.99	0.90	0.97	9.97	3.98	0.14	39.98	58.01
10%	2.21	2.24	19.42	6.97	8.02	15.00	2.81	2.83	9.13	9.29	0.24	42.11	3.06	3.08	9.13	9.94	0.21	42.54	39.30
15%	4.64	4.65	18.14	13.32	1.36	23.72	5.49	5.50	8.20	15.67	0.53	42.47	5.68	5.69	8.05	17.92	0.21	44.23	27.84
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.46	0.56	20.00	3.09	0.28	14.96	0.63	0.72	10.18	3.09	0.26	39.32	0.95	1.01	10.12	4.38	0.14	40.38	57.63
10%	2.09	2.12	19.39	7.06	8.04	15.43	2.69	2.71	9.28	9.67	0.23	42.52	2.69	2.71	9.28	9.67	0.23	42.52	39.24
15%	4.19	4.21	18.71	13.37	1.31	24.32	4.63	4.64	8.88	14.47	0.57	42.39	4.97	4.98	8.37	17.81	0.21	44.71	27.92
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.92	0.97	19.79	2.64	0.26	15.10	-0.72	0.79	9.94	2.64	0.24	39.88	-0.85	0.92	9.91	3.86	0.14	40.81	56.89
10%	-1.48	1.52	18.79	5.78	8.22	15.92	-1.56	1.60	8.82	8.69	0.24	42.83	-1.56	1.60	8.82	8.69	0.24	42.83	36.66
15%	-2.38	2.40	18.22	12.64	1.35	25.66	-2.36	2.39	8.62	13.97	0.60	43.21	-2.93	2.95	8.17	18.11	0.23	45.97	25.15
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-1.07	1.11	20.08	2.66	8.23	10.51	-0.90	0.96	10.13	2.66	0.24	39.26	-1.09	1.14	10.07	3.87	0.14	40.25	56.84
10%	-1.68	1.72	19.03	6.28	8.01	16.04	-1.94	1.97	9.01	8.97	0.24	42.87	-1.94	1.97	9.01	8.97	0.24	42.87	37.51
15%	-2.88	2.90	18.14	12.69	1.30	25.44	-3.04	3.06	8.65	13.95	0.56	42.74	-3.43	3.45	8.22	17.62	0.22	45.23	26.29

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.05	0.23	1.29	1.01	0.23	4.72	-0.04	0.23	1.27	1.09	0.21	4.85	-0.01	0.22	1.24	2.06	0.07	5.89	54.95
10%	-0.12	0.26	1.68	4.24	0.35	12.97	-0.09	0.25	1.59	4.75	0.16	13.71	-0.05	0.23	1.22	4.75	0.14	15.19	31.72
15%	-0.01	0.23	1.66	9.47	0.46	27.20	-0.01	0.23	1.66	9.47	0.46	27.20	-0.08	0.25	1.38	11.83	0.12	29.92	17.84
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.67	0.71	1.50	1.73	0.28	4.56	0.74	0.77	1.48	1.89	0.15	4.78	1.02	1.05	1.44	3.06	0.10	5.91	58.17
10%	2.66	2.67	1.79	6.74	0.45	10.01	2.92	2.93	1.70	7.55	0.22	10.90	3.24	3.25	1.33	8.23	0.18	13.01	39.79
15%	5.05	5.06	1.82	13.23	0.68	20.29	5.53	5.53	1.68	14.42	0.40	21.60	5.96	5.97	1.55	15.74	0.15	23.26	28.01
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.61	0.65	1.45	1.84	0.27	4.80	0.66	0.70	1.43	2.02	0.15	5.03	0.94	0.97	1.39	3.12	0.09	6.11	58.34
10%	2.48	2.50	1.77	6.92	0.45	9.71	2.71	2.72	1.69	7.61	0.21	10.51	2.95	2.96	1.36	8.28	0.17	12.58	39.67
15%	4.41	4.41	1.89	13.05	0.68	19.39	4.67	4.68	1.73	14.12	0.40	20.64	5.02	5.02	1.62	15.37	0.15	22.22	28.26
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.25	0.33	1.46	1.25	0.25	5.10	-0.25	0.34	1.44	1.38	0.23	5.28	-0.34	0.40	1.39	2.52	0.07	6.47	56.36
10%	-0.75	0.78	1.74	5.81	0.38	12.11	-0.83	0.86	1.66	6.52	0.19	12.95	-0.81	0.84	1.33	6.52	0.16	14.46	37.02
15%	-1.40	1.42	1.84	12.02	0.58	23.75	-1.40	1.42	1.78	12.02	0.38	24.15	-1.87	1.89	1.59	14.68	0.15	26.70	25.37
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.28	0.36	1.44	1.38	0.26	4.80	-0.30	0.38	1.42	1.57	0.24	5.01	-0.40	0.46	1.39	2.71	0.08	6.18	56.99
10%	-1.06	1.08	1.83	5.94	0.40	12.02	-1.18	1.20	1.76	6.63	0.19	12.87	-1.15	1.17	1.37	6.63	0.16	14.52	37.78
15%	-2.24	2.25	1.92	12.38	0.58	24.48	-2.24	2.25	1.86	12.38	0.36	24.91	-2.70	2.71	1.65	14.94	0.14	27.42	26.31

Table S.A.17: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets with of 100,000 records each. In addition, there is a 50% overlap and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields To declare a pair of records to be in the set of matches, three different thresholds are considered. For each linkage field, missing comparisons have been coded as disagreements.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
MISSING COMPLETELY AT RANDOM (MCAR)																			
5%	-0.02	0.18	0.18	1.02	0.17	3.20	-0.02	0.18	0.18	1.02	0.15	3.21	0.02	0.18	0.17	2.12	0.07	4.37	54.57
10%	-0.00	0.18	0.25	3.68	0.40	8.03	-0.03	0.18	0.18	4.59	0.10	10.05	-0.03	0.18	0.18	4.59	0.10	10.06	31.56
15%	-0.02	0.19	0.31	8.21	0.35	19.57	-0.04	0.19	0.27	9.42	0.31	20.69	-0.02	0.19	0.23	11.62	0.08	22.98	17.91
MISSING NOT AT RANDOM (MNAR): INDEPENDENCE																			
5%	0.67	0.70	0.17	1.69	0.22	3.00	0.75	0.77	0.17	1.87	0.18	3.19	1.09	1.11	0.16	3.09	0.09	4.40	58.02
10%	2.70	2.70	0.23	6.88	0.37	6.91	2.99	3.00	0.18	7.62	0.14	8.60	2.99	3.00	0.18	7.62	0.11	8.67	39.32
15%	5.17	5.18	0.25	13.21	0.49	13.58	5.18	5.18	0.25	13.21	0.30	13.82	6.00	6.00	0.19	15.56	0.10	16.85	27.78
MISSING NOT AT RANDOM (MNAR): DEPENDENCE																			
5%	0.63	0.66	0.19	1.70	0.22	3.01	0.70	0.73	0.18	1.89	0.18	3.19	1.01	1.02	0.18	3.06	0.09	4.39	58.07
10%	2.43	2.44	0.24	6.86	0.35	7.01	2.68	2.69	0.19	7.58	0.14	8.64	2.68	2.69	0.19	7.58	0.11	8.70	39.55
15%	4.49	4.50	0.25	13.29	0.50	13.79	4.49	4.50	0.24	13.29	0.31	14.02	5.20	5.20	0.19	15.66	0.10	17.11	28.05
MISSING AT RANDOM (MAR): INDEPENDENCE																			
5%	-0.15	0.23	0.17	1.27	0.19	3.09	-0.16	0.24	0.16	1.38	0.17	3.21	-0.24	0.30	0.16	2.54	0.08	4.38	56.46
10%	-0.66	0.69	0.26	5.04	0.47	7.24	-0.80	0.82	0.18	6.30	0.13	9.53	-0.80	0.82	0.18	6.30	0.10	9.60	36.79
15%	-1.32	1.33	0.35	10.70	0.48	16.10	-1.48	1.49	0.31	11.99	0.42	17.32	-1.80	1.81	0.21	14.68	0.09	20.79	25.41
MISSING AT RANDOM (MAR): DEPENDENCE																			
5%	-0.18	0.25	0.19	1.40	0.20	3.09	-0.20	0.27	0.19	1.56	0.17	3.25	-0.29	0.34	0.18	2.75	0.08	4.46	57.00
10%	-0.95	0.97	0.23	5.94	0.33	7.94	-1.07	1.09	0.18	6.66	0.13	9.60	-1.07	1.09	0.18	6.66	0.10	9.67	37.53
15%	-2.08	2.09	0.32	12.42	0.46	16.21	-2.08	2.09	0.30	12.43	0.28	16.59	-2.56	2.56	0.22	14.99	0.09	19.70	26.32

Table S.A.18: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets with of 100,000 records each. In addition, there is a 80% overlap and five missing data mechanisms are studied across three different proportions of missing data for each linkage field other than year of birth. Measurement error is introduced to all the string valued fields To declare a pair of records to be in the set of matches, three different thresholds are considered. For each linkage field, missing comparisons have been coded as disagreements.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-1.96	3.92	10.55	2.50	0.42	9.96	-2.19	4.05	10.14	2.50	0.22	10.94	-2.20	4.06	9.77	3.00	0.11	12.38	34.50
50%	-2.40	4.13	13.06	3.50	0.65	28.36	-2.40	4.13	13.06	3.50	0.65	28.36	-1.73	3.85	9.39	3.50	0.08	33.44	10.50
80%	-1.26	3.69	14.83	11.00	1.85	33.24	-1.99	4.02	14.08	11.50	0.76	36.13	-2.01	4.09	11.11	12.00	0.23	41.24	5.00
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.83	3.44	12.00	1.00	0.32	12.32	-0.38	3.38	11.21	1.00	0.14	13.28	0.32	3.39	10.00	1.00	0.07	14.36	37.00
50%	-1.07	3.47	15.65	3.00	2.35	28.11	-1.95	3.87	13.84	3.50	1.16	35.32	-1.50	3.75	11.79	6.50	0.16	42.20	16.00
80%	-3.19	4.72	14.98	12.00	2.64	29.98	-3.74	5.14	12.44	12.00	1.39	33.60	-3.58	5.09	9.90	13.50	0.15	41.31	6.00
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-1.50	3.67	12.95	2.50	0.58	19.06	-1.27	3.60	12.67	3.50	0.30	20.28	-1.05	3.54	12.27	3.50	0.07	22.22	34.50
50%	-0.83	3.46	16.14	6.5	2.15	24.61	-2.19	4.05	13.82	6.50	1.15	33.81	-1.74	3.89	10.63	7.50	0.15	40.55	15.50
80%	-1.02	3.61	12.86	8.5	2.07	28.13	-0.03	3.51	10.78	9.00	0.60	32.04	1.05	3.73	8.67	10.50	0.15	36.11	11.00
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.61	3.40	13.39	3.00	0.56	11.56	-0.61	3.40	13.39	3.00	0.50	11.85	-1.27	3.65	11.16	4.50	0.08	15.30	35.50
50%	-1.50	3.75	13.68	8.50	1.63	24.85	-2.22	4.12	12.92	9.00	0.90	26.71	-2.26	4.23	10.66	12.00	0.15	32.87	16.50
80%	-3.13	4.63	15.81	9.50	2.55	25.06	-2.72	4.43	12.68	10.50	0.66	32.58	-3.28	4.85	9.64	11.00	0.19	37.13	8.00
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-1.96	3.92	12.39	4.5	0.31	10.10	-1.73	3.82	11.98	4.5	0.24	11.05	-1.50	3.74	11.21	5.00	0.06	12.45	37.00
50%	-3.90	5.12	17.03	5.0	1.59	28.27	-3.30	4.71	15.32	6.00	0.68	32.63	-1.98	3.98	11.90	7.50	0.14	36.41	22.50
80%	-4.41	5.62	14.08	11.5	1.77	32.59	-4.47	5.69	13.86	13.0	1.19	34.11	-3.83	5.26	10.88	14.00	0.13	41.35	7.50

Table S.A.19: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string–valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered.

Table S.A.20 presents the following results across three thresholds.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-0.49	2.27	4.70	2.60	0.26	5.98	-0.59	2.29	4.71	2.80	0.20	6.41	-0.80	2.37	3.59	3.20	0.07	8.25	35.20
50%	0.00	2.24	4.20	4.20	0.47	6.71	0.20	2.26	3.83	4.60	0.16	7.49	0.71	2.37	3.46	5.20	0.08	9.45	14.60
80%	-0.10	2.28	4.95	7.80	1.63	15.36	0.00	2.28	4.38	8.20	0.18	20.27	0.11	2.30	3.38	8.60	0.08	22.40	4.40
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.29	2.23	3.52	1.40	0.14	4.15	-0.29	2.23	3.52	1.40	0.14	4.14	0.00	2.22	3.35	1.80	0.05	5.12	34.60
50%	0.90	2.41	5.39	5.20	0.33	9.72	1.10	2.50	5.21	5.40	0.24	10.01	1.42	2.67	4.07	5.60	0.05	12.22	16.80
80%	2.80	3.64	4.09	11.00	0.43	11.08	3.06	3.85	4.15	12.20	0.30	11.94	3.42	4.15	3.75	12.80	0.09	14.33	9.00
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.20	2.23	4.15	3.00	0.34	3.92	0.20	2.24	3.77	3.00	0.23	4.38	0.20	2.26	3.23	4.00	0.02	7.06	35.60
50%	1.39	2.63	4.56	3.80	0.73	5.77	1.11	2.50	4.02	4.60	0.23	7.17	1.54	2.74	2.87	5.40	0.09	9.70	19.20
80%	1.23	2.58	4.30	6.60	0.76	11.46	1.47	2.72	3.78	8.40	0.22	14.01	1.49	2.75	2.98	8.80	0.08	15.34	9.60
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.10	2.22	3.72	1.60	0.17	5.71	-0.10	2.22	3.72	1.60	0.11	5.86	0.20	2.23	3.54	2.00	0.06	6.49	37.00
50%	-0.30	2.25	5.35	4.40	0.52	10.39	-0.80	2.38	4.82	5.20	0.10	12.00	-0.71	2.37	3.46	5.20	0.06	13.33	16.60
80%	-0.80	2.38	7.23	7.60	2.50	13.28	-0.84	2.44	5.44	9.60	0.58	19.99	-0.75	2.43	3.84	9.80	0.13	27.05	7.80
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.78	2.35	3.32	1.00	0.26	3.60	-0.59	2.29	2.94	1.00	0.14	4.20	-0.69	2.33	2.76	1.40	0.02	5.55	32.40
50%	-1.38	2.62	6.13	5.00	0.40	13.43	-1.49	2.68	5.94	5.00	0.26	13.86	-1.71	2.82	4.63	5.20	0.08	16.10	16.40
80%	-0.73	2.40	5.20	8.80	0.56	12.95	-1.17	2.58	4.67	10.20	0.28	14.44	-1.29	2.65	4.08	10.60	0.12	15.65	8.80

Table S.A.20: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string–valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	0.37	1.82	0.38	1.88	0.10	1.77	0.37	1.82	0.38	1.88	0.08	1.87	0.37	1.82	0.38	2.12	0.01	2.77	33.88
50%	-0.14	1.79	0.89	2.88	0.55	3.10	-0.21	1.80	0.64	3.00	0.16	4.51	-0.21	1.80	0.64	3.25	0.07	5.37	14.00
80%	-0.11	1.83	0.67	7.00	0.45	6.93	-0.18	1.84	0.67	7.12	0.27	7.81	0.15	1.85	0.41	8.75	0.05	10.65	4.00
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.37	1.82	0.64	2.38	0.10	2.37	0.37	1.82	0.64	2.38	0.09	2.47	0.31	1.81	0.64	2.50	0.03	3.39	35.88
50%	0.75	1.94	0.77	2.75	0.39	5.47	0.62	1.90	0.64	3.12	0.22	6.22	0.50	1.87	0.52	3.75	0.06	7.58	18.12
80%	2.01	2.74	1.39	11.00	0.46	8.13	2.15	2.85	1.39	11.25	0.35	8.56	2.47	3.11	1.14	13.00	0.08	11.06	7.00
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.43	1.84	0.64	2.50	0.06	2.26	0.43	1.84	0.64	2.50	0.06	2.26	0.43	1.84	0.64	2.50	0.04	2.49	36.88
50%	1.08	2.10	0.91	4.25	0.23	3.78	1.34	2.25	0.91	4.75	0.12	4.64	1.28	2.21	0.91	5.12	0.03	5.78	20.38
80%	0.76	1.99	1.08	8.00	0.27	7.49	0.96	2.07	1.08	8.38	0.21	7.77	0.97	2.08	0.95	9.00	0.05	9.72	6.88
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.20	1.79	1.01	2.12	0.26	2.17	-0.14	1.79	0.89	2.38	0.18	2.66	-0.27	1.80	0.89	2.62	0.06	3.85	34.25
50%	-0.02	1.80	1.03	4.00	0.42	4.75	0.17	1.81	0.91	4.25	0.16	5.37	-0.03	1.81	0.65	4.62	0.08	7.02	16.62
80%	-1.08	2.14	0.95	9.00	0.35	7.83	-1.08	2.14	0.95	9.00	0.19	8.17	-1.17	2.19	0.83	10.00	0.06	9.75	7.38
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.06	1.78	0.76	1.62	0.17	2.77	-0.01	1.78	0.63	1.62	0.08	3.16	-0.07	1.78	0.51	1.88	0.06	3.48	36.75
50%	-0.41	1.84	1.03	4.25	0.42	4.78	-0.67	1.92	0.78	4.50	0.17	5.88	-0.95	2.05	0.52	5.25	0.07	7.18	18.88
80%	-1.34	2.27	0.95	8.50	0.50	7.27	-1.27	2.24	0.95	8.62	0.45	7.46	-1.39	2.32	0.83	10.62	0.06	10.50	8.25

Table S.A.21: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string–valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-0.85	1.36	10.64	2.20	0.46	11.58	-0.87	1.38	10.52	2.20	0.43	11.72	-0.63	1.25	8.54	2.60	0.07	15.08	33.60
50%	-0.90	1.40	11.40	4.05	1.07	17.46	-0.78	1.33	10.25	4.15	0.57	19.79	-0.63	1.26	8.37	4.70	0.14	22.84	15.30
80%	-1.01	1.50	10.87	8.95	1.13	24.14	-1.19	1.63	9.86	9.00	0.59	25.53	-1.11	1.59	7.49	9.80	0.14	30.01	3.75
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.30	1.11	10.95	3.20	0.68	8.66	-0.42	1.16	9.95	3.20	0.30	10.26	-0.06	1.09	8.59	3.70	0.07	12.52	34.75
50%	-0.04	1.07	11.28	4.05	1.14	15.34	0.01	1.08	10.37	4.10	0.70	17.07	0.50	1.21	8.71	5.20	0.19	20.49	18.40
80%	1.24	1.66	11.69	9.70	0.71	22.59	1.23	1.66	11.41	9.90	0.61	22.98	1.22	1.66	8.70	10.85	0.14	27.44	7.20
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.55	1.20	10.57	2.25	0.63	10.51	-0.59	1.23	9.46	2.35	0.29	12.19	-0.21	1.11	7.97	2.95	0.07	14.63	36.00
50%	-0.42	1.15	12.45	4.70	1.14	17.22	-0.42	1.16	11.84	5.05	0.71	19.03	-0.18	1.11	10.10	6.10	0.14	22.97	18.95
80%	0.82	1.37	10.94	8.80	0.77	21.01	0.78	1.35	10.74	8.95	0.69	21.33	0.96	1.48	7.89	10.05	0.16	26.21	8.25
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-1.02	1.48	10.32	2.20	0.59	9.42	-0.98	1.45	9.07	2.25	0.23	11.16	-0.79	1.34	8.11	2.55	0.06	12.73	33.60
50%	-1.59	1.92	11.91	4.25	1.08	19.76	-1.58	1.91	11.60	4.70	0.88	20.49	-1.55	1.89	9.27	5.55	0.14	25.77	17.45
80%	-1.92	2.21	12.10	9.55	0.71	24.63	-1.87	2.17	11.72	9.60	0.61	24.97	-1.55	1.92	8.91	10.60	0.14	29.60	7.40
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-1.18	1.59	10.12	2.30	0.65	10.10	-1.25	1.65	8.95	2.30	0.27	11.88	-0.96	1.45	7.66	2.90	0.05	14.18	36.65
50%	-1.46	1.81	12.64	4.95	1.08	16.65	-1.49	1.84	11.87	5.35	0.73	18.08	-1.66	1.99	10.14	6.50	0.14	22.11	19.10
80%	-3.15	3.34	11.60	10.80	1.11	24.61	-3.13	3.32	10.69	11.00	0.59	26.07	-3.12	3.32	8.39	12.05	0.13	30.91	7.90

Table S.A.22: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior $\geq$ 0.75						Threshold 2: Posterior $\geq$ 0.85						Threshold 3: Posterior $\geq$ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-0.01	0.71	2.21	1.78	0.23	4.24	-0.06	0.71	2.11	1.82	0.15	4.54	-0.02	0.71	1.73	2.26	0.04	5.80	34.12
50%	-0.25	0.75	3.15	3.46	0.59	5.78	-0.32	0.78	3.05	4.02	0.17	6.74	-0.21	0.74	2.49	4.36	0.10	8.49	15.46
80%	-0.21	0.75	3.15	6.66	0.72	11.09	-0.28	0.78	2.94	7.68	0.32	12.68	-0.36	0.81	2.67	8.30	0.12	14.44	4.56
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.23	0.74	2.53	2.18	0.24	3.80	0.25	0.75	2.43	2.20	0.16	4.00	0.29	0.77	1.98	2.82	0.05	5.32	34.98
50%	0.59	0.92	3.31	4.32	0.74	7.05	0.76	1.04	2.90	5.04	0.32	8.73	0.92	1.17	2.50	5.52	0.10	10.56	18.50
80%	1.81	1.95	3.10	8.70	0.49	10.51	1.97	2.10	2.84	9.70	0.32	11.65	2.03	2.16	2.12	10.40	0.09	13.81	6.80
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.19	0.73	2.64	2.50	0.40	3.88	0.23	0.74	2.31	2.54	0.17	4.65	0.40	0.81	2.00	3.12	0.05	6.04	36.34
50%	0.59	0.93	3.02	4.30	0.70	7.24	0.64	0.96	2.71	5.08	0.35	8.59	0.77	1.05	2.18	5.60	0.10	10.70	18.60
80%	1.10	1.32	3.04	8.74	0.62	9.76	1.25	1.45	2.82	9.78	0.27	11.02	1.40	1.59	2.29	10.24	0.09	12.95	7.22
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.03	0.71	2.74	2.20	0.50	3.49	-0.01	0.71	2.24	2.22	0.15	4.76	-0.02	0.71	1.80	2.66	0.04	6.00	34.66
50%	-0.15	0.72	3.26	3.76	0.72	7.05	-0.24	0.75	2.93	4.42	0.34	8.41	-0.17	0.74	2.28	4.90	0.10	10.43	16.18
80%	-0.78	1.07	3.66	8.36	0.84	11.76	-0.85	1.12	3.43	9.34	0.36	13.40	-0.75	1.05	2.91	9.86	0.12	15.50	7.40
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.00	0.70	2.70	1.92	0.51	3.21	0.01	0.71	2.22	2.00	0.18	4.31	0.10	0.72	1.85	2.54	0.04	5.64	34.88
50%	-0.68	0.98	3.25	4.24	0.75	6.47	-0.72	1.02	2.91	5.26	0.33	8.16	-0.75	1.04	2.40	5.78	0.10	10.05	17.84
80%	-0.86	1.12	3.31	8.80	0.53	9.38	-0.94	1.19	3.12	9.88	0.32	10.63	-0.89	1.16	2.33	10.42	0.09	12.69	7.80

Table S.A.23: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-0.05	0.56	0.51	1.88	0.17	2.09	-0.04	0.56	0.46	1.94	0.10	2.39	-0.01	0.56	0.43	2.11	0.06	2.75	35.30
50%	0.15	0.58	0.82	2.96	0.50	3.61	0.14	0.58	0.68	3.40	0.16	4.89	0.15	0.59	0.54	3.79	0.06	6.00	15.06
80%	-0.06	0.58	0.77	6.06	0.40	7.16	-0.12	0.59	0.68	6.38	0.28	7.62	-0.12	0.59	0.43	7.45	0.06	9.72	4.16
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.31	0.64	0.61	2.29	0.18	2.21	0.36	0.67	0.57	2.39	0.12	2.59	0.40	0.69	0.47	2.58	0.06	3.09	35.88
50%	0.82	1.00	0.89	3.80	0.47	4.16	0.98	1.13	0.73	4.28	0.15	5.65	1.13	1.27	0.60	4.67	0.06	6.75	18.71
80%	1.41	1.52	0.85	7.85	0.42	7.55	1.46	1.57	0.77	8.25	0.32	8.24	1.58	1.68	0.59	9.50	0.06	10.43	8.14
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.29	0.63	0.56	2.08	0.23	2.36	0.33	0.65	0.47	2.14	0.12	2.83	0.35	0.67	0.46	2.42	0.06	3.32	36.76
50%	0.73	0.92	0.90	3.84	0.51	3.91	0.87	1.04	0.73	4.33	0.19	5.57	0.96	1.11	0.60	4.85	0.07	6.80	18.88
80%	1.33	1.45	1.01	8.12	0.64	7.76	1.35	1.47	0.96	8.53	0.38	8.59	1.60	1.71	0.68	9.89	0.07	11.20	7.25
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.00	0.56	0.68	1.99	0.20	2.16	0.01	0.56	0.62	2.10	0.11	2.50	0.03	0.56	0.55	2.24	0.06	2.93	34.95
50%	-0.11	0.58	0.86	3.58	0.39	4.27	-0.03	0.57	0.71	3.86	0.15	5.45	-0.04	0.57	0.60	4.25	0.06	6.45	17.34
80%	-0.70	0.91	0.99	7.81	0.40	8.01	-0.61	0.84	0.92	8.12	0.30	8.66	-0.66	0.88	0.69	9.41	0.06	10.95	7.75
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.11	0.57	0.58	1.81	0.22	2.22	-0.10	0.57	0.52	1.88	0.11	2.67	-0.11	0.57	0.47	2.10	0.06	3.12	35.91
50%	-0.30	0.64	0.86	3.95	0.43	4.46	-0.34	0.67	0.80	4.49	0.16	5.14	-0.43	0.71	0.61	5.03	0.08	6.63	18.21
80%	-1.03	1.18	1.08	8.06	0.44	7.80	-1.08	1.23	0.95	8.40	0.36	8.39	-1.13	1.27	0.65	9.92	0.06	10.98	7.74

Table S.A.24: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered.

Table S.A.25 presents the following results across three thresholds.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-0.23	0.42	7.43	2.42	0.57	9.43	-0.18	0.39	6.80	2.47	0.24	10.97	-0.21	0.41	6.01	2.91	0.07	12.78	34.57
50%	-0.40	0.52	8.83	4.06	1.09	15.89	-0.35	0.49	7.99	4.25	0.59	17.89	-0.22	0.41	6.53	5.03	0.13	21.42	15.05
80%	-0.21	0.41	8.29	8.44	1.12	22.28	-0.21	0.41	7.53	8.67	0.59	23.79	-0.15	0.39	5.64	9.58	0.13	28.43	4.14
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.12	0.36	7.80	2.36	0.68	9.05	-0.04	0.35	7.14	2.43	0.28	10.79	0.05	0.35	6.20	2.90	0.07	12.76	35.70
50%	0.73	0.81	9.06	5.10	1.14	14.37	0.81	0.88	8.21	5.42	0.67	16.24	1.11	1.16	6.91	6.25	0.16	19.74	18.34
80%	1.83	1.87	8.54	10.11	0.74	23.72	1.90	1.93	8.25	10.37	0.62	24.17	2.12	2.15	6.36	11.42	0.14	28.70	7.33
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.03	0.34	7.92	2.33	0.64	9.14	0.04	0.35	7.25	2.41	0.28	10.82	0.15	0.38	6.41	2.89	0.07	12.81	35.51
50%	0.78	0.86	9.32	5.05	1.17	14.97	0.82	0.89	8.58	5.27	0.70	16.84	0.99	1.05	7.17	6.31	0.17	20.39	18.64
80%	1.23	1.28	8.73	10.47	0.75	21.89	1.26	1.31	8.48	10.73	0.64	22.29	1.49	1.54	6.62	11.81	0.15	26.85	7.75
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.29	0.45	7.68	2.09	0.65	9.09	-0.27	0.44	6.93	2.17	0.25	10.90	-0.27	0.44	6.12	2.59	0.07	12.82	35.45
50%	-0.56	0.66	9.12	4.54	1.10	15.97	-0.56	0.66	8.22	4.72	0.66	17.81	-0.56	0.66	6.72	5.61	0.14	21.54	16.59
80%	-1.17	1.22	8.79	9.77	0.73	22.40	-1.16	1.22	8.60	10.05	0.61	22.81	-1.17	1.22	6.76	11.09	0.14	27.29	7.62
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.41	0.53	7.68	2.36	0.59	9.63	-0.35	0.49	7.03	2.42	0.24	11.27	-0.37	0.51	6.19	2.87	0.07	13.15	35.19
50%	-0.70	0.78	9.26	4.81	1.07	16.91	-0.77	0.84	8.57	5.11	0.68	18.68	-0.83	0.90	7.04	6.06	0.15	22.39	18.00
80%	-1.36	1.40	8.66	10.48	0.77	22.81	-1.38	1.42	8.46	10.73	0.66	23.20	-1.35	1.39	6.73	11.94	0.16	27.87	7.97

Table S.A.25: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets of 100,000 records each. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-0.07	0.23	1.27	2.09	0.23	3.78	-0.07	0.23	1.20	2.11	0.18	3.97	-0.05	0.23	0.98	2.58	0.05	5.22	34.68
50%	-0.02	0.23	1.53	3.49	0.49	5.97	-0.06	0.23	1.43	3.95	0.35	6.68	-0.05	0.23	1.13	4.28	0.10	8.55	14.77
80%	-0.07	0.24	1.54	6.69	0.68	9.80	-0.05	0.24	1.35	7.63	0.28	11.31	-0.03	0.23	1.13	8.28	0.13	12.68	4.35
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.20	0.30	1.36	2.09	0.38	3.75	0.24	0.33	1.21	2.15	0.17	4.54	0.25	0.34	0.98	2.56	0.04	5.79	35.76
50%	0.93	0.96	1.66	4.48	0.65	5.99	1.04	1.06	1.47	5.05	0.33	7.15	1.16	1.18	1.17	5.57	0.10	8.97	18.55
80%	1.61	1.62	1.84	8.56	0.75	9.99	1.82	1.83	1.63	9.56	0.30	11.59	1.85	1.87	1.26	10.12	0.08	13.83	7.73
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.17	0.28	1.39	2.13	0.38	3.48	0.22	0.31	1.23	2.21	0.17	4.21	0.26	0.34	1.00	2.63	0.05	5.41	35.68
50%	0.73	0.77	1.64	4.42	0.62	6.03	0.81	0.84	1.50	4.96	0.33	7.07	0.92	0.95	1.21	5.40	0.10	8.88	18.65
80%	1.36	1.38	1.78	8.65	0.56	9.73	1.49	1.51	1.62	9.62	0.31	11.23	1.52	1.54	1.23	10.14	0.08	13.47	7.97
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.10	0.25	1.34	2.00	0.26	3.53	-0.09	0.24	1.18	2.02	0.18	4.26	-0.09	0.24	0.94	2.35	0.04	5.48	34.84
50%	-0.21	0.31	1.64	3.69	0.59	6.25	-0.24	0.33	1.49	4.22	0.31	7.33	-0.25	0.34	1.20	4.65	0.09	9.19	16.74
80%	-0.63	0.67	1.76	8.08	0.56	10.34	-0.71	0.75	1.60	9.11	0.31	11.81	-0.71	0.75	1.41	9.71	0.08	13.39	7.67
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.14	0.26	1.37	2.17	0.39	3.40	-0.13	0.26	1.20	2.22	0.16	4.17	-0.15	0.27	0.99	2.57	0.05	5.24	35.24
50%	-0.47	0.52	1.68	4.13	0.63	6.17	-0.51	0.56	1.52	4.67	0.33	7.31	-0.53	0.58	1.21	5.15	0.10	9.22	18.47
80%	-1.14	1.16	1.82	8.79	0.58	9.98	-1.29	1.31	1.63	9.84	0.31	11.55	-1.34	1.36	1.28	10.44	0.08	13.75	8.23

Table S.A.26: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets of 100,000 records each. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior $\geq$ 0.75						Threshold 2: Posterior $\geq$ 0.85						Threshold 3: Posterior $\geq$ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	0.06	0.19	0.16	2.01	0.20	2.38	0.06	0.19	0.13	2.08	0.10	2.82	0.06	0.19	0.12	2.23	0.06	3.20	34.59
50%	0.03	0.18	0.20	3.08	0.45	3.80	0.05	0.19	0.18	3.57	0.15	5.08	0.03	0.18	0.15	3.91	0.07	6.03	15.03
80%	-0.03	0.19	0.24	6.46	0.46	7.81	-0.03	0.19	0.22	6.82	0.33	8.57	-0.07	0.20	0.18	8.24	0.09	10.57	4.38
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.25	0.30	0.15	2.13	0.19	2.45	0.27	0.33	0.14	2.18	0.11	2.79	0.31	0.35	0.12	2.40	0.06	3.26	35.93
50%	0.88	0.90	0.20	4.10	0.46	4.51	1.03	1.05	0.18	4.57	0.28	5.21	1.16	1.18	0.17	5.10	0.09	6.69	18.27
80%	1.77	1.78	0.30	8.06	0.62	7.93	1.81	1.81	0.27	8.47	0.36	8.75	2.05	2.06	0.20	9.88	0.07	11.30	7.63
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.26	0.32	0.16	2.11	0.21	2.49	0.27	0.33	0.14	2.15	0.12	2.90	0.32	0.37	0.13	2.39	0.06	3.43	35.73
50%	0.85	0.87	0.20	4.18	0.45	4.27	0.99	1.01	0.16	4.69	0.15	5.69	1.08	1.10	0.15	5.26	0.06	6.58	18.76
80%	1.37	1.38	0.27	8.24	0.49	7.61	1.40	1.41	0.22	8.61	0.34	8.45	1.55	1.57	0.17	9.97	0.07	10.80	7.92
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.02	0.18	0.15	2.04	0.19	2.40	0.02	0.18	0.13	2.11	0.10	2.81	0.02	0.18	0.11	2.28	0.06	3.17	35.13
50%	-0.10	0.20	0.22	3.42	0.42	4.18	-0.14	0.23	0.17	3.84	0.15	5.50	-0.18	0.25	0.16	4.32	0.08	6.32	16.81
80%	-0.45	0.49	0.30	7.66	0.51	7.71	-0.47	0.51	0.27	8.02	0.33	8.49	-0.56	0.59	0.19	9.45	0.07	10.97	7.68
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.03	0.18	0.16	2.04	0.19	2.35	-0.04	0.18	0.14	2.12	0.11	2.77	-0.04	0.18	0.13	2.32	0.06	3.19	35.24
50%	-0.32	0.37	0.24	3.91	0.45	4.33	-0.38	0.42	0.22	4.39	0.28	5.00	-0.41	0.45	0.18	4.89	0.08	6.43	18.46
80%	-0.89	0.91	0.31	8.38	0.51	7.84	-0.90	0.92	0.27	8.75	0.34	8.57	-1.04	1.06	0.18	10.15	0.07	11.07	8.24

Table S.A.27: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets of 100,000 records each. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-1.72	3.76	13.33	2.50	1.16	21.08	-2.18	4.03	10.96	2.50	0.52	23.26	-1.27	3.65	9.30	2.50	0.23	25.47	34.50
50%	-0.61	3.40	13.39	3.00	1.40	48.28	-1.50	3.73	10.65	3.50	0.49	50.25	-0.54	3.52	8.17	4.50	0.19	52.13	10.50
80%	-1.26	3.67	13.27	8.50	1.50	58.82	-0.52	3.55	11.76	10.00	0.82	60.10	-0.47	3.63	9.28	12.00	0.39	61.83	4.50
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.39	3.36	12.00	1.00	0.48	22.35	0.07	3.35	11.21	1.00	0.34	22.89	0.08	3.37	10.86	1.50	0.21	24.12	37.00
50%	-1.29	3.49	18.41	2.50	1.83	45.78	-1.29	3.53	16.31	2.50	1.12	47.67	-0.08	3.44	12.26	7.00	0.34	52.62	1.50
80%	-2.01	4.09	11.11	12.00	1.06	45.79	-2.27	4.25	9.74	12.00	0.63	46.86	-2.56	4.46	7.45	13.00	0.24	49.32	6.00
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.83	3.44	13.33	2.50	0.63	35.06	-1.28	3.59	12.56	2.50	0.50	35.53	-1.04	3.55	11.93	4.00	0.23	37.56	34.50
50%	-1.92	3.77	100.00	100.00	14.37	80.83	-1.50	5.51	100.00	100.00		100.00	-0.52	3.55	10.78	9.00	0.19	46.17	15.50
80%	-0.55	3.50	13.33	9.00	1.04	48.65	-0.30	3.48	12.92	9.00	0.68	49.75	1.32	3.82	10.77	13.00	0.24	53.06	11.00
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.20	3.31	14.78	2.00	1.07	25.40	-0.15	3.37	12.61	3.00	0.34	27.70	0.33	3.41	11.47	3.50	0.23	28.88	35.50
50%	-2.17	4.00	17.04	7.50	1.31	46.21	-2.64	4.29	15.98	8.00	0.83	47.35	-3.21	4.75	12.68	10.50	0.18	51.46	16.50
80%	-2.18	4.03	15.98	8.00	1.55	56.08	-2.68	4.37	13.74	9.00	0.77	57.94	-2.74	4.47	10.45	10.00	0.32	60.66	8.00
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-2.58	4.19	16.88	4.00	1.25	26.87	-1.04	3.55	12.39	4.50	0.34	30.57	-1.27	3.66	10.80	5.00	0.14	32.57	37.00
50%	-2.19	4.05	11.52	4.00	0.97	40.23	-2.20	4.08	10.33	4.50	0.60	41.47	-2.47	4.28	10.19	7.50	0.33	43.31	3.50
80%	-2.94	4.55	12.98	9.50	1.82	46.35	-3.00	4.64	11.50	11.50	1.13	48.55	-2.29	4.29	7.85	12.00	0.43	52.39	0.50

Table S.A.28: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered. Conditional independence across linkage fields is not assumed and a log-linear model is used to model dependencies among non-matches.

29

Table S.A.29

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-0.49	2.27	4.52	2.80	0.32	11.00	-0.49	2.28	4.34	3.00	0.26	11.28	-0.80	2.37	3.59	3.20	0.10	12.84	35.20
50%	0.00	2.22	5.93	4.80	0.48	19.86	0.40	2.24	5.18	4.80	0.28	20.68	0.40	2.28	4.44	5.20	0.16	21.52	14.60
80%	0.51	2.32	5.50	7.20	0.95	31.02	0.32	2.28	3.95	7.60	0.36	32.81	0.32	2.33	2.76	8.40	0.16	34.39	4.40
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.19	2.21	4.46	1.40	0.39	7.90	0.10	2.21	3.91	1.60	0.12	8.90	0.10	2.22	3.54	1.80	0.08	9.84	34.60
50%	1.60	2.75	5.20	5.20	0.47	19.28	1.95	2.77	4.44	5.20	0.26	20.04	1.95	2.99	3.70	6.20	0.13	21.49	16.80
80%	3.07	3.83	5.07	10.20	0.92	28.26	3.20	4.01	4.13	11.80	0.31	30.24	3.20	3.97	3.53	12.60	0.14	31.72	2.40
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.49	2.27	5.09	3.00	0.65	11.15	0.40	2.31	4.18	3.80	0.22	12.95	0.40	2.28	3.43	4.20	0.11	14.19	35.60
50%	1.89	2.92	4.57	4.00	0.83	17.46	1.98	3.01	3.61	4.00	0.55	18.67	1.98	3.03	2.30	6.40	0.08	22.25	19.20
80%	1.41	2.65	4.44	5.20	0.86	24.12	1.70	2.85	3.89	6.00	0.60	25.20	1.70	2.87	2.77	8.60	0.19	27.90	9.60
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.20	2.22	3.91	1.60	0.33	9.80	0.10	5.78	100.00	100.00	13.16	98.64	0.10	2.23	3.17	2.20	0.09	11.61	37.00
50%	-0.20	2.24	5.16	4.40	0.80	17.41	-0.61	2.39	4.05	5.20	0.24	19.56	-0.61	2.34	3.86	5.40	0.20	19.99	16.60
80%	0.00	2.25	6.65	7.40	0.67	34.10	-0.10	2.26	5.89	7.40	0.66	36.19	-0.10	2.29	5.24	9.60	0.20	38.77	2.20
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.49	2.26	3.88	1.00	0.39	6.43	-0.50	2.32	3.52	1.40	0.21	7.26	-0.50	2.29	1.99	1.40	0.05	9.59	32.40
50%	-1.29	2.58	5.17	4.60	0.47	24.15	-1.41	2.55	4.62	5.00	0.28	24.86	-1.41	2.65	4.64	5.40	0.16	26.07	16.40
80%	0.32	2.32	2.96	8.20	0.76	27.71	-0.88	2.31	2.77	8.80	0.51	28.41	-0.88	2.51	1.55	11.00	0.19	31.22	8.80

Table S.A.29: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered. Conditional independence across linkage fields is not assumed and a log-linear model is used to model dependencies among non-matches.

	Threshold 1: Posterior $\geq$ 0.75						Threshold 2: Posterior $\geq$ 0.85						Threshold 3: Posterior $\geq$ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	0.37	1.82	0.38	1.88	0.17	4.39	0.31	1.81	0.25	1.88	0.09	4.79	0.24	1.81	0.26	2.50	0.04	5.33	33.88
50%	-0.34	1.82	0.51	2.88	0.28	11.26	-0.27	1.81	0.51	3.00	0.14	11.79	-0.34	1.82	0.39	3.00	0.08	12.30	14.00
80%	-0.30	1.85	0.93	6.62	0.49	19.51	-0.11	1.83	0.80	7.12	0.36	20.14	0.08	1.85	0.41	8.62	0.08	22.41	4.00
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.62	1.88	2.01	2.25	9.38	8.98	0.50	1.85	1.02	2.50	0.21	39.74	0.62	1.90	0.64	3.12	0.06	41.13	35.88
50%	0.69	1.91	0.76	2.62	0.47	9.95	0.56	1.88	0.64	3.00	0.29	10.67	0.62	1.90	0.52	3.75	0.13	12.07	18.12
80%	1.99	2.72	1.10	9.75	0.80	19.13	2.08	2.79	1.11	10.62	0.48	20.35	2.68	3.28	0.85	12.62	0.13	23.72	2.62
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.37	1.82	0.76	2.50	0.13	4.85	0.43	1.84	0.76	2.62	0.09	5.05	0.50	1.85	0.77	2.75	0.07	5.37	36.88
50%	1.01	2.06	0.90	4.12	0.38	11.32	1.21	2.17	0.78	4.38	0.15	12.31	1.34	2.25	0.78	5.12	0.08	13.15	3.50
80%	1.09	2.13	1.60	7.62	0.70	22.02	1.03	2.11	1.21	8.38	0.25	23.63	1.03	2.11	1.09	9.00	0.11	24.64	2.00
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.13	1.78	1.14	2.12	0.30	6.09	-0.13	1.78	1.01	2.25	0.22	6.55	-0.27	1.81	0.77	2.75	0.07	8.04	34.25
50%	-0.02	1.79	1.16	3.88	0.55	11.13	0.04	1.80	0.91	4.25	0.19	12.47	-0.09	1.81	0.78	4.62	0.11	13.17	16.62
80%	-1.20	2.20	1.22	8.75	0.61	17.24	-1.07	2.13	1.09	8.88	0.32	18.29	-1.09	2.15	0.96	10.00	0.11	19.93	1.75
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.13	1.78	0.51	1.75	0.15	7.47	-0.13	1.78	0.51	1.75	0.13	7.56	-0.07	1.78	0.51	2.12	0.06	8.28	36.75
50%	-0.54	1.88	0.78	4.25	0.39	11.79	-0.48	1.86	0.78	4.38	0.24	12.43	-0.95	2.05	0.66	5.62	0.12	13.65	18.88
80%	-1.33	2.27	1.08	8.38	0.58	16.82	-1.41	2.32	0.95	8.62	0.48	17.34	-1.39	2.32	0.83	10.62	0.12	19.90	8.25

Table S.A.30: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 1,000 records. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered. Conditional independence across linkage fields is not assumed and a log-linear model is used to model dependencies among non-matches.

	Threshold 1: Posterior $\geq$ 0.75						Threshold 2: Posterior $\geq$ 0.85						Threshold 3: Posterior $\geq$ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	0.41	1.15	8.82	2.25	0.79	6.82	-0.41	1.15	8.82	2.25	0.79	6.82	-0.10	1.12	3.00	3.10	0.09	13.76	33.60
50%	-0.60	1.25	7.59	4.40	0.93	9.46	-0.60	1.25	7.59	4.40	0.93	9.46	-0.39	1.20	2.71	6.55	0.15	16.82	15.30
80%	-1.06	1.54	7.84	8.90	0.80	16.12	-1.06	1.54	7.84	8.90	0.80	16.11	-0.88	1.45	5.61	10.85	0.38	19.87	3.05
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.08	1.09	8.27	3.45	0.37	5.59	0.08	1.09	8.27	3.45	0.37	5.59	0.65	1.30	2.20	4.30	0.10	11.88	34.75
50%	0.62	1.26	7.99	4.40	0.91	10.75	0.62	1.26	7.99	4.40	0.91	10.74	1.92	2.22	6.35	7.75	0.16	15.70	3.20
80%	1.11	1.58	8.68	9.55	1.00	18.71	1.11	1.58	8.68	9.55	1.00	18.71	1.25	1.70	5.56	11.60	0.41	23.79	2.10
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.11	1.09	7.51	2.65	0.95	5.67	-0.11	1.09	7.51	2.65	0.95	5.67	0.30	1.16	2.87	3.60	0.13	11.70	36.00
50%	0.10	1.10	9.33	5.30	1.11	10.53	0.10	1.10	9.33	5.30	1.11	10.53	0.97	1.50	3.23	8.50	0.15	19.87	2.85
80%	1.07	1.55	7.39	8.55	1.03	15.96	1.07	1.55	7.39	8.55	1.03	15.97	1.94	2.26	3.17	11.40	0.34	22.74	1.80
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.53	1.21	7.13	2.35	1.32	5.93	-0.53	1.21	7.13	2.35	1.32	5.93	-0.20	1.14	2.02	2.85	0.09	11.80	33.60
50%	-1.02	1.49	8.22	4.55	0.79	12.37	-1.02	1.49	8.22	4.55	0.79	12.36	-1.24	1.67	6.43	6.80	0.29	16.29	17.45
80%	-1.65	1.99	9.02	9.20	1.15	16.60	-1.65	1.99	9.02	9.20	1.15	16.61	-0.81	1.42	3.35	12.10	0.33	24.53	2.00
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-1.01	1.49	7.37	2.60	0.41	4.96	-1.01	1.49	7.37	2.60	0.41	4.96	-0.31	1.17	2.34	3.90	0.09	11.18	36.65
50%	-1.42	1.79	8.84	5.70	0.75	13.21	-1.42	1.79	8.84	5.70	0.75	13.20	-2.22	2.49	8.15	8.75	0.24	16.95	19.10
80%	-2.73	2.95	8.09	10.80	1.21	15.67	-2.73	2.95	8.09	10.80	1.21	15.68	-2.29	2.57	3.39	13.00	0.36	22.57	1.55

Table S.A.31: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered. Conditional independence across linkage fields is not assumed and a log-linear model is used to model dependencies among non-matches.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	0.07	0.71	0.83	1.92	0.24	3.11	0.07	0.72	0.51	2.10	0.13	3.64	0.13	0.73	0.35	2.58	0.05	4.44	34.12
50%	-0.20	0.74	1.88	3.88	0.59	4.77	-0.22	0.75	1.13	4.12	0.26	6.13	-0.08	0.73	0.90	5.06	0.12	7.41	3.66
80%	0.00	0.72	1.97	6.58	0.83	8.24	-0.18	0.75	1.33	7.66	0.39	10.05	-0.21	0.77	0.87	8.90	0.16	11.98	2.70
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.37	0.80	0.75	2.36	0.19	2.87	0.37	0.80	0.55	2.54	0.14	3.28	0.49	0.87	0.54	3.36	0.06	4.13	2.00
50%	0.87	1.13	0.96	4.84	0.48	5.84	1.02	1.25	0.76	5.40	0.25	6.69	1.05	1.28	0.51	6.28	0.11	7.98	2.80
80%	1.88	2.01	2.14	8.68	0.70	10.22	2.04	2.17	1.94	9.86	0.51	11.39	2.20	2.32	0.85	11.14	0.16	14.24	6.80
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.31	0.78	0.92	2.56	0.35	2.83	0.43	0.84	0.68	2.94	0.13	3.54	0.57	0.92	0.35	3.64	0.05	4.60	36.34
50%	0.73	1.02	2.14	4.94	0.82	5.70	0.86	1.12	1.62	5.42	0.50	6.85	1.19	1.40	0.66	6.74	0.16	9.73	2.64
80%	1.13	1.34	1.68	8.62	0.82	9.85	1.40	1.58	1.03	9.74	0.34	11.71	1.58	1.75	0.69	11.04	0.17	13.39	1.56
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.06	0.72	0.47	2.38	0.18	1.99	0.09	0.72	0.41	2.58	0.13	2.22	0.08	0.72	0.10	3.10	0.04	3.28	34.66
50%	-0.10	0.72	2.10	4.14	0.71	5.74	-0.08	0.72	1.77	4.72	0.41	6.97	-0.19	0.75	0.86	5.74	0.14	9.49	2.88
80%	-0.66	0.98	2.32	8.30	0.93	10.42	-0.77	1.06	1.71	9.34	0.42	12.32	-0.73	1.04	0.80	10.54	0.16	14.77	1.60
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.00	0.71	0.67	2.12	0.27	2.81	0.03	0.71	0.43	2.36	0.12	3.36	0.03	0.72	0.35	2.94	0.06	4.00	34.88
50%	-0.69	1.00	2.14	5.06	0.73	5.71	-0.67	0.98	1.73	5.46	0.28	6.81	-0.78	1.07	0.85	6.34	0.15	8.99	2.70
80%	-0.75	1.05	2.19	8.74	0.91	10.12	-0.85	1.12	1.64	9.82	0.46	12.05	-0.98	1.23	0.74	11.18	0.16	14.63	1.90

Table S.A.32: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered. Conditional independence across linkage fields is not assumed and a log-linear model is used to model dependencies among non-matches.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-0.02	0.56	0.43	1.95	0.36	2.01	0.02	0.56	0.22	2.03	0.11	3.01	0.03	0.57	0.20	2.49	0.05	3.64	35.30
50%	0.17	0.59	0.59	3.01	0.55	3.60	0.22	0.61	0.31	3.62	0.17	5.02	0.19	0.60	0.23	4.11	0.09	5.79	15.06
80%	-0.02	0.57	0.42	6.08	0.33	6.73	-0.08	0.58	0.33	6.39	0.29	7.13	-0.11	0.59	0.22	7.62	0.12	8.61	4.16
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.38	0.68	0.23	2.33	0.16	1.06	0.46	0.73	0.13	2.55	0.07	1.51	0.59	0.82	0.05	2.99	0.04	2.02	35.88
50%	0.91	1.07	0.58	3.76	0.61	3.54	1.17	1.30	0.30	4.42	0.21	5.01	1.32	1.44	0.22	5.20	0.10	6.22	2.96
80%	1.50	1.61	0.30	7.84	0.45	7.98	1.55	1.66	0.24	8.20	0.34	8.42	1.76	1.85	0.17	9.72	0.12	10.21	2.25
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.42	0.70	0.43	2.33	0.44	1.98	0.43	0.71	0.19	2.44	0.13	2.99	0.51	0.77	0.15	3.05	0.05	3.80	36.76
50%	0.76	0.95	0.58	3.81	0.63	3.72	1.01	1.16	0.43	4.66	0.24	5.42	1.18	1.31	0.18	5.33	0.12	6.55	3.20
80%	1.42	1.54	0.32	7.99	0.51	7.64	1.39	1.51	0.27	8.39	0.35	8.27	1.78	1.87	0.15	10.14	0.13	10.31	2.02
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.04	0.56	0.17	2.01	0.22	0.94	0.01	0.56	0.08	2.21	0.09	1.42	0.03	0.57	0.04	2.54	0.04	1.90	34.95
50%	-0.07	0.57	0.44	3.53	0.46	4.11	-0.05	0.57	0.36	3.99	0.21	4.94	-0.12	0.58	0.24	4.72	0.11	6.04	4.61
80%	-0.63	0.86	0.56	7.58	0.61	7.68	-0.62	0.85	0.39	8.06	0.36	8.61	-0.65	0.87	0.23	9.62	0.13	10.67	2.00
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.09	0.57	0.20	1.88	0.22	1.34	-0.17	0.59	0.05	2.10	0.08	1.99	-0.21	0.60	0.03	2.47	0.04	2.49	4.01
50%	-0.31	0.65	0.40	4.06	0.42	4.28	-0.33	0.66	0.34	4.69	0.23	5.04	-0.44	0.72	0.24	5.34	0.12	5.95	2.88
80%	-1.02	1.17	0.51	7.96	0.54	7.66	-1.10	1.24	0.41	8.35	0.38	8.32	-1.30	1.43	0.28	10.22	0.12	10.55	1.66

Table S.A.33: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets where one has 100,000 records and the other one has 10,000 records. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered. Conditional independence across linkage fields is not assumed and a log-linear model is used to model dependencies among non-matches.

	Threshold 1: Posterior $\geq$ 0.75						Threshold 2: Posterior $\geq$ 0.85						Threshold 3: Posterior $\geq$ 0.95						
	Bias	RMSE	FDR	$\widehat{FDR}$	FNR	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	-0.06	0.36	0.39	0.19	2.77	2.45	-0.06	0.36	0.38	2.80	0.18	2.48	-0.07	0.37	0.17	3.72	0.06	3.60	4.21
50%	-0.02	0.36	1.51	0.56	4.17	4.60	0.01	0.36	1.41	4.67	0.43	5.15	0.06	0.37	0.42	6.45	0.13	8.13	3.50
80%	-0.03	0.37	2.38	0.85	8.09	11.79	-0.02	0.37	1.86	8.70	0.46	12.95	0.01	0.37	1.22	10.83	0.25	15.59	2.44
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.34	0.49	0.48	0.24	2.71	2.82	0.40	0.53	0.43	2.86	0.16	3.03	0.53	0.64	0.29	3.62	0.09	3.90	3.65
50%	1.20	1.25	1.54	0.62	5.49	5.32	1.38	1.43	1.46	5.94	0.47	5.82	2.17	2.20	0.44	8.56	0.13	9.57	2.96
80%	2.02	2.06	2.08	0.86	9.77	11.66	2.20	2.23	1.57	10.58	0.51	13.03	2.59	2.62	1.15	12.99	0.27	15.72	1.85
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.39	0.53	0.56	0.23	2.72	2.75	0.41	0.54	0.50	2.82	0.20	2.89	0.47	0.59	0.27	3.58	0.09	3.85	3.98
50%	1.10	1.16	1.46	0.68	5.33	5.21	1.26	1.31	1.35	5.92	0.55	5.83	1.99	2.02	0.34	8.73	0.15	9.70	2.65
80%	1.48	1.53	2.51	0.87	10.25	13.37	1.64	1.68	1.93	11.31	0.51	14.94	1.99	2.03	1.32	13.47	0.26	17.60	1.60
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.00	0.36	0.44	0.21	2.37	2.45	0.00	0.36	0.42	2.44	0.18	2.55	-0.04	0.36	0.24	3.20	0.07	3.49	2.42
50%	-0.11	0.37	1.70	0.59	4.72	6.40	-0.17	0.40	1.41	5.19	0.42	7.13	-0.28	0.46	1.18	7.52	0.14	9.68	3.12
80%	-0.75	0.83	1.80	0.77	9.43	10.33	-0.81	0.89	1.54	9.89	0.59	11.02	-0.89	0.96	0.86	11.97	0.26	13.91	1.54
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.17	0.40	0.43	0.20	2.77	2.42	-0.18	0.40	0.43	2.78	0.20	2.45	-0.23	0.43	0.18	3.36	0.08	3.34	3.77
50%	-0.47	0.60	1.53	0.63	5.10	5.26	-0.54	0.65	1.41	5.62	0.51	5.83	-0.98	1.04	0.55	8.08	0.15	9.30	1.32
80%	-1.08	1.14	2.56	0.96	10.19	13.82	-1.13	1.19	2.12	10.90	0.58	14.95	-1.33	1.38	1.25	13.02	0.27	18.03	1.65

Table S.A.34: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets of 100,000 records each. In addition, there is a 20% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered. Conditional independence across linkage fields is not assumed and a log-linear model is used to model dependencies among non-matches.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						Exact
	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	Bias	RMSE	FDR	FNR	$\widehat{\text{FDR}}$	$\widehat{\text{FNR}}$	
CLASSICAL MEASUREMENT ERROR																			
20%	-0.01	0.23	0.08	2.18	0.14	1.27	-0.02	0.23	0.07	2.43	0.09	1.50	-0.02	0.23	0.05	2.95	0.04	2.04	2.43
50%	-0.02	0.23	0.25	3.82	0.31	2.72	-0.05	0.23	0.22	4.15	0.22	3.08	-0.08	0.24	0.08	5.01	0.10	4.44	3.63
80%	-0.04	0.23	0.33	6.58	0.54	5.62	-0.04	0.24	0.23	7.59	0.29	6.66	-0.01	0.23	0.17	8.72	0.17	7.92	2.52
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.36	0.43	0.11	2.32	0.15	1.48	0.36	0.43	0.08	2.47	0.11	1.66	0.47	0.52	0.05	3.11	0.04	2.34	2.05
50%	1.11	1.13	0.31	4.85	0.44	2.90	1.26	1.28	0.24	5.46	0.28	3.57	1.44	1.46	0.10	6.28	0.14	4.91	2.79
80%	1.67	1.69	0.39	8.43	0.61	6.13	1.87	1.88	0.29	9.51	0.33	7.27	2.01	2.02	0.18	10.73	0.17	8.84	1.60
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.31	0.38	0.10	2.38	0.14	1.51	0.29	0.37	0.08	2.49	0.10	1.67	0.42	0.48	0.05	3.15	0.04	2.36	2.18
50%	0.90	0.93	0.29	4.77	0.45	3.01	1.03	1.06	0.23	5.39	0.23	3.71	1.16	1.18	0.09	6.20	0.12	5.00	2.69
80%	1.40	1.42	0.38	8.51	0.59	6.46	1.52	1.54	0.29	9.56	0.32	7.61	1.60	1.62	0.16	10.75	0.17	9.09	1.64
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	-0.05	0.23	0.10	2.14	0.15	1.39	-0.07	0.24	0.07	2.36	0.09	1.65	-0.10	0.25	0.05	2.95	0.04	2.27	2.31
50%	-0.19	0.30	0.33	3.93	0.38	2.78	-0.20	0.30	0.27	4.40	0.24	3.34	-0.23	0.33	0.11	5.26	0.12	4.63	3.03
80%	-0.57	0.62	0.39	7.97	0.59	6.43	-0.66	0.70	0.28	9.11	0.31	7.63	-0.72	0.76	0.17	10.47	0.17	9.25	1.69
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.12	0.26	0.09	2.38	0.15	1.43	-0.13	0.26	0.08	2.52	0.11	1.60	-0.19	0.29	0.05	3.12	0.04	2.23	2.16
50%	-0.42	0.48	0.29	4.43	0.39	2.95	-0.47	0.52	0.22	5.05	0.27	3.59	-0.57	0.61	0.07	5.98	0.14	5.02	2.80
80%	-1.07	1.09	0.42	8.64	0.63	6.38	-1.25	1.27	0.30	9.82	0.35	7.62	-1.43	1.45	0.17	11.15	0.18	9.27	1.49

Table S.A.35: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{\text{FDR}}$), the estimated false negative rate ($\widehat{\text{FNR}}$), and the rate of exact matches (exact), when merging two datasets of 100,000 records each. In addition, there is a 50% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered. Conditional independence across linkage fields is not assumed and a log-linear model is used to model dependencies among non-matches.

	Threshold 1: Posterior ≥ 0.75						Threshold 2: Posterior ≥ 0.85						Threshold 3: Posterior ≥ 0.95						
	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Bias	RMSE	FDR	FNR	$\widehat{FDR}$	$\widehat{FNR}$	Exact
CLASSICAL MEASUREMENT ERROR																			
20%	0.07	0.19	0.01	2.08	0.12	0.93	0.06	0.19	0.01	2.16	0.09	1.04	0.06	0.19	0.01	2.53	0.04	1.45	2.46
50%	0.03	0.18	0.05	3.10	0.35	1.89	0.06	0.19	0.04	3.69	0.17	2.53	0.04	0.18	0.02	4.29	0.09	3.40	3.56
80%	-0.04	0.19	0.08	6.38	0.46	4.99	-0.03	0.19	0.06	6.76	0.34	5.44	-0.09	0.20	0.03	8.75	0.12	7.53	2.58
DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.29	0.34	0.02	2.20	0.15	1.11	0.37	0.41	0.02	2.39	0.07	1.32	0.43	0.47	0.01	2.79	0.04	1.72	2.13
50%	0.90	0.92	0.05	4.08	0.43	2.20	1.13	1.15	0.04	4.82	0.22	2.97	1.39	1.40	0.02	5.70	0.11	4.12	2.79
80%	1.79	1.80	0.08	7.97	0.51	5.82	1.83	1.84	0.06	8.39	0.36	6.34	2.17	2.17	0.03	10.15	0.12	8.28	1.67
DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	0.31	0.36	0.02	2.20	0.15	1.06	0.38	0.42	0.02	2.43	0.07	1.29	0.44	0.47	0.01	2.84	0.04	1.74	2.12
50%	0.87	0.89	0.05	4.17	0.41	2.29	1.09	1.10	0.04	5.06	0.22	3.19	1.26	1.28	0.01	5.88	0.11	4.28	2.81
80%	1.37	1.39	0.07	8.14	0.51	5.65	1.41	1.42	0.06	8.53	0.37	6.09	1.66	1.67	0.02	10.25	0.13	8.04	1.63
NON-DIFFERENTIAL MEASUREMENT ERROR: INDEPENDENCE																			
20%	0.03	0.18	0.02	2.11	0.12	1.08	0.02	0.18	0.02	2.22	0.09	1.21	-0.00	0.18	0.01	2.60	0.04	1.63	2.28
50%	-0.09	0.20	0.05	3.42	0.38	2.17	-0.16	0.24	0.03	4.18	0.18	2.98	-0.20	0.27	0.02	4.84	0.10	3.89	2.92
80%	-0.42	0.46	0.08	7.56	0.48	5.65	-0.44	0.48	0.07	7.94	0.36	6.08	-0.59	0.62	0.03	9.72	0.14	8.02	1.64
NON-DIFFERENTIAL MEASUREMENT ERROR: DEPENDENCE																			
20%	-0.04	0.18	0.02	2.14	0.14	1.02	-0.06	0.19	0.02	2.29	0.09	1.19	-0.09	0.20	0.01	2.71	0.04	1.62	2.20
50%	-0.30	0.35	0.05	3.90	0.40	2.31	-0.38	0.42	0.04	4.64	0.21	3.08	-0.49	0.52	0.02	5.41	0.11	4.07	2.71
80%	-0.87	0.89	0.10	8.29	0.50	6.02	-0.88	0.90	0.09	8.69	0.37	6.47	-1.09	1.11	0.03	10.77	0.13	8.77	1.57

Table S.A.36: The table presents the estimation error (Bias) for the turnout rate, root mean square error (RMSE), the false discovery rate (FDR), the false negative rate (FNR), the estimated false discovery rate ($\widehat{FDR}$), the estimated false negative rate ($\widehat{FNR}$), and the rate of exact matches (exact), when merging two datasets of 100,000 records each. In addition, there is a 80% overlap (measured as a percentage of the smaller data set) and five measurement error mechanisms are studied across three different proportions of missing data for the following string-valued linkage fields: first name, last name, and street name. To declare a pair of records to be in the set of matches, three different thresholds are considered. Conditional independence across linkage fields is not assumed and a log-linear model is used to model dependencies among non-matches.