

## Supplementary Appendices

**Supplemental Table 1.** Agreement conditions, match ( $m$ ) probabilities, proportion collected, and proportion of records with agreement for each field ( $i$ ) in the probabilistic algorithm, by first and matched search attempts

**Supplemental Figure 1.** Log frequency of match scores calculated for all pairwise comparisons using full algorithm, by true match status

**Supplemental Figure 2.** Log frequency of match scores calculated for all pairwise comparisons using limited algorithm, by true match status

**Supplemental Figure 3.** Sensitivity ( $Se$ ) and positive predictive value (PPV) of automated retrospective record linkage at various match score percentile thresholds, full (F) vs. limited (L) algorithm

**Supplemental Table 2.** Distribution of individual characteristics in the dataset matched using point-of-contact interactive record linkage (PIRL) compared with those matched using a purely automated probabilistic approach using a full and limited algorithm, by match score threshold

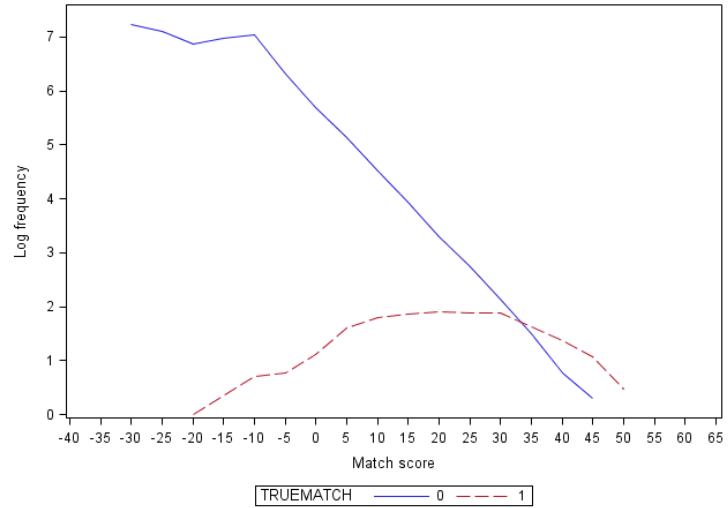
Supplemental Table 1: Agreement conditions, match (m) probabilities, proportion collected, and proportion of records with agreement for each field (i) in the probabilistic algorithm, by first and matched search attempts,  $n_M=2,612$

Field <i>i</i>	Agreement condition	m-prob	First search		Matched search		Change ( $\Delta$ )=matched-first	
			% collected	% agreement	% collected	% agreement	$\Delta$ % collected	$\Delta$ % agreement
First name	Jaro-Winkler $\geq 0.8$	0.87	100.0%	83.8%	100.0%	94.1%	0.0%	10.3%
Second name	Jaro-Winkler $\geq 0.8$	0.87	100.0%	77.9%	100.0%	87.9%	0.0%	10.1%
Third name	Jaro-Winkler $\geq 0.8$	0.85	83.4%	5.7%	82.0%	5.3%	-1.4%	-0.3%
TCL first name	Jaro-Winkler $\geq 0.8$	0.87	44.8%	15.1%	65.8%	42.9%	20.9%	27.8%
TCL second name	Jaro-Winkler $\geq 0.8$	0.87	39.4%	13.6%	60.8%	40.9%	21.5%	27.3%
TCL third name	Jaro-Winkler $\geq 0.8$	0.85	0.2%	0.0%	0.2%	0.2%	0.0%	0.1%
HH first name	Jaro-Winkler $\geq 0.8$	0.52	90.5%	70.1%	93.2%	75.2%	2.7%	5.1%
HH second name	Jaro-Winkler $\geq 0.8$	0.52	89.6%	64.3%	92.2%	70.8%	2.6%	6.5%
HH third name	Jaro-Winkler $\geq 0.8$	0.52	4.1%	1.1%	4.4%	1.1%	0.3%	0.0%
Sex	exact match	0.99	99.8%	97.6%	99.8%	97.7%	0.0%	0.1%
Year of birth	within 2 years	0.80	98.7%	84.9%	99.1%	87.0%	0.4%	2.1%
Month of birth	exact match	0.63	3.7%	1.4%	4.0%	1.6%	0.3%	0.2%
Day of birth	exact match	0.57	3.6%	1.0%	3.9%	1.2%	0.3%	0.2%
Village	exact match	0.89	90.9%	83.3%	93.0%	89.4%	2.1%	6.1%
Sub-village	exact match	0.89	90.9%	67.2%	93.0%	78.0%	2.1%	10.8%

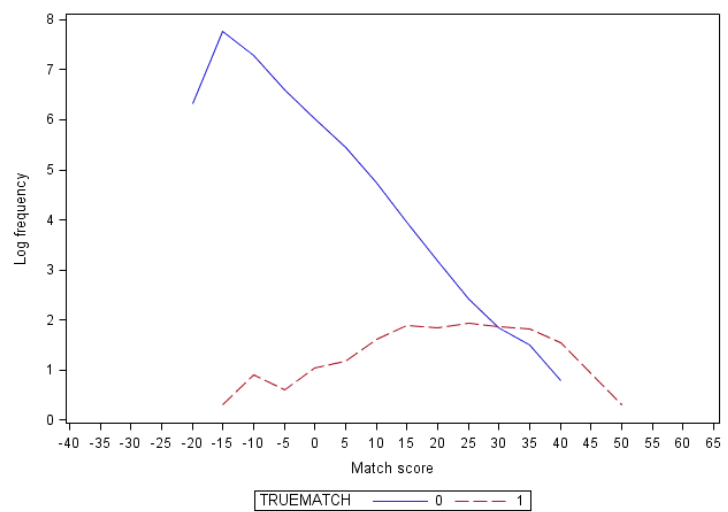
Abbreviations: HDSS = health and demographic surveillance surveys;  $n_M$  = number of matches; m-prob = match probability; TCL = ten-cell leader; HH = household member

Notes: TCL = an individual for a group of ten households; % collected = proportion of matched records with completed information; % agreement = proportion of matched records with agreeing information

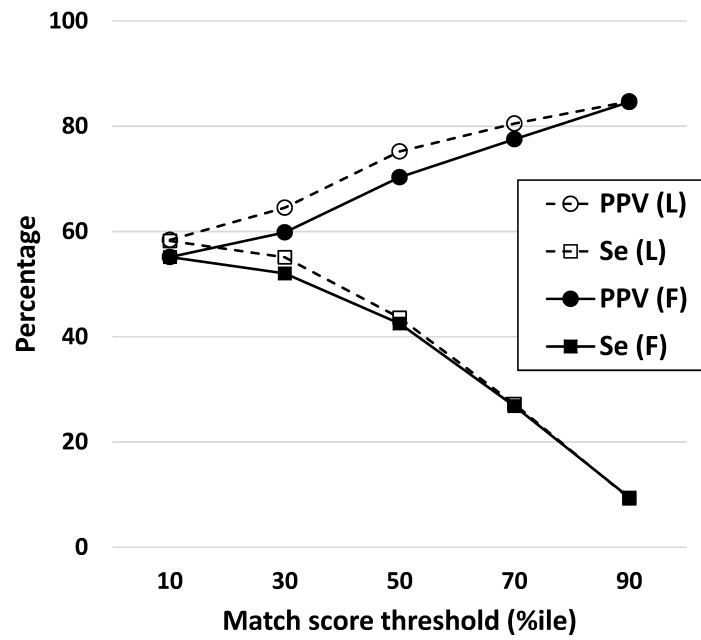
Supplemental Figure 1: Log frequency of match scores calculated for all pairwise comparisons using full algorithm, by true match status



Supplemental Figure 2: Log frequency of match scores calculated for all pairwise comparisons using limited algorithm, by true match status



Supplemental Figure 3: Sensitivity (Se) and positive predictive value (PPV) of automated retrospective record linkage at various match score percentile thresholds, full (F) vs. limited (L) algorithm



Supplemental Table 2: Distribution of individual characteristics in the dataset matched using point-of-contact interactive record linkage (PIRL) compared with those matched using a purely automated probabilistic approach using a full and limited algorithm, by match score threshold

Characteristic	PIRL match n (%)	Automated: full algorithm						Automated: limited algorithm					
		Threshold=10%ile		Threshold=50%ile		Threshold=90%ile		Threshold=10%ile		Threshold=50%ile		Threshold=90%ile	
		n (%)	p-value*	n (%)	p-value*	n (%)	p-value*	n (%)	p-value*	n (%)	p-value*	n (%)	p-value*
Total matched (PPV)	2,612	2,612 (55.1)		1,579 (70.3)		292 (84.6)		2602 (58.4)		1,514 (75.2)		288 (84.7)	
Sex													
<i>Female</i>	2,061 (78.9)	2,036 (78.0)	0.4004	1,185 (75.1)	0.0038	213 (73.0)	0.0191	2,059 (79.1)	0.8409	1,158 (76.5)	0.0706	209 (72.6)	0.0133
<i>Male</i>	551 (21.1)	576 (22.1)		394 (25.0)		79 (27.1)		543 (20.9)		356 (23.5)		79 (27.4)	
Age, in years													
<5	125 (4.8)	198 (7.6)	<0.0001	132 (8.4)	<0.0001	46 (15.8)	<0.0001	198 (7.6)	<0.0001	122 (8.1)	0.0013	33 (11.5)	<0.0001
5-17	393 (15.1)	464 (17.8)		239 (15.2)		35 (12.0)		453 (17.4)		211 (14.0)		34 (11.8)	
18-34	1,384 (53.0)	1,301 (49.9)		770 (48.8)		125 (42.8)		1,325 (51.0)		765 (50.6)		121 (42.0)	
35-49	522 (20.0)	433 (16.6)		301 (19.1)		68 (23.3)		437 (16.8)		296 (19.6)		74 (25.7)	
50-64	160 (6.1)	162 (6.2)		105 (6.7)		15 (5.1)		144 (5.5)		99 (6.5)		23 (8.0)	
65+	28 (1.1)	52 (2.0)		30 (1.9)		3 (1.0)		43 (1.7)		20 (1.3)		3 (1.0)	
Village of residence													
<i>Kisesa</i>	999 (38.3)	982 (37.6)	0.9340	586 (37.1)	0.8100	111 (38.0)	0.3320	981 (37.7)	0.6773	531 (35.1)	0.3071	73 (25.4)	0.0002
<i>Kanyama</i>	521 (20.0)	529 (20.3)		302 (19.1)		46 (15.8)		527 (20.3)		299 (19.8)		59 (20.5)	
<i>Kitumba</i>	424 (16.2)	444 (17.0)		262 (16.6)		48 (16.4)		436 (16.8)		254 (16.8)		49 (17.0)	
<i>Isangijo</i>	257 (9.8)	258 (9.9)		176 (11.2)		39 (13.4)		254 (9.8)		177 (11.7)		46 (16.0)	
<i>Ihayabuyaga</i>	152 (5.8)	138 (5.3)		89 (5.6)		21 (7.2)		129 (5.0)		87 (5.8)		22 (7.6)	
<i>Igekemaja</i>	141 (5.4)	150 (5.7)		94 (6.0)		13 (4.5)		163 (6.3)		94 (6.2)		24 (8.3)	
<i>Welamasonga</i>	118 (4.5)	111 (4.3)		70 (4.4)		14 (4.8)		112 (4.3)		72 (4.8)		15 (5.2)	
Marital status <sup>a</sup>													
<i>Never married</i>	362 (24.0)	272 (24.1)	0.9997	179 (22.5)	0.4266	33 (22.3)	0.6089	286 (25.3)	0.8668	176 (22.5)	0.7093	26 (16.5)	0.0139
<i>Married once</i>	724 (48.0)	540 (47.8)		403 (50.6)		72 (48.7)		536 (47.4)		391 (49.9)		80 (50.6)	
<i>Remarried</i>	175 (11.6)	132 (11.7)		99 (12.4)		22 (14.9)		124 (11.0)		95 (12.1)		30 (19.0)	
<i>Separated/Widowed</i>	249 (16.5)	187 (16.5)		116 (14.6)		21 (14.2)		185 (16.4)		121 (15.5)		22 (13.9)	
Pregnant at last HDSS round <sup>b</sup>													
<i>No</i>	1,057 (95.7)	758 (95.5)	0.8425	529 (95.0)	0.5292	101 (98.1)	0.3094	769 (95.2)	0.6166	531 (95.5)	0.8862	93 (92.1)	0.1310
<i>Yes</i>	48 (4.3)	36 (4.5)		28 (5.0)		2 (1.9)		39 (4.8)		25 (4.5)		8 (7.9)	
Enrolled in school at last HDSS round <sup>c</sup>													
<i>No</i>	378 (72.0)	282 (67.6)	0.1454	185 (68.3)	0.2725	25 (52.1)	0.0038	295 (67.7)	0.1438	186 (69.9)	0.5422	21 (60.0)	0.1288
<i>Yes</i>	147 (28.0)	135 (32.4)		86 (31.7)		23 (47.9)		141 (32.3)		80 (30.1)		14 (40.0)	

Abbreviations: HDSS - health and demographic sentinel surveillance

\*Statistical differences tested for significance with chi-square ( $\chi^2$ ) or Fisher's Exact tests

<sup>a</sup>This question was only given to individuals aged 15 years or older

<sup>b</sup>This question was only given to females between 15 and 49 years of age

<sup>c</sup>This question was only given to individuals between 5 and 25 years of age