

SPONSORED BY



Date Range Propagation in Genealogical Databases

Randy Wilson

wilsonr@familysearch.org

Family History Technology Workshop
(FHTW 2012)



Different snapshots

Robert Jones, b. 1820

Bob Jones, m. 1860 to Mary Lee

Rob Jones, d. 1810

Inferring Missing Data

Robert Jones, b. 1820

=> m. 1835..1890; d. 1820-1917

Bob Jones, m. 1860 to Mary Lee

=> b. 1790..1845; d. 1860-1930

Rob Jones, d. 1810

=> b. 1720..1810; m. 1740..1810

Uses for date propagation

- Matching
 - *Are these the same real person?*
- Searching
 - *Which results are reasonable?*
- Living calculation
 - *Could this person still be alive?*

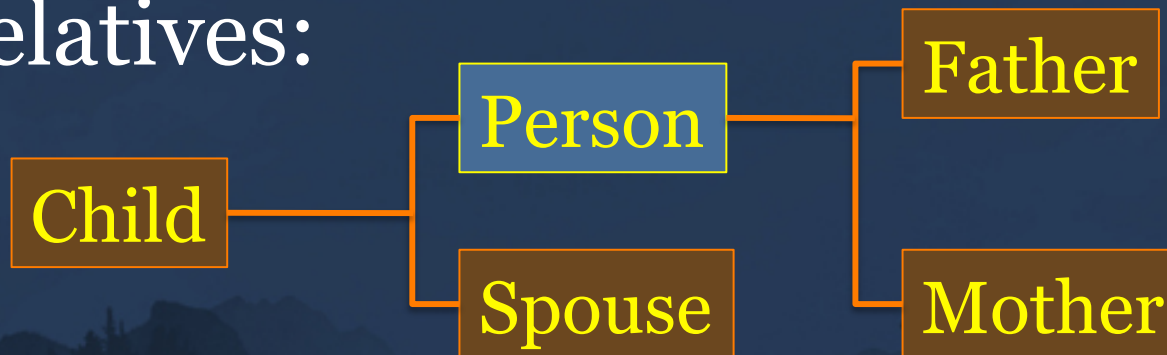
Problem definition

G = Relationship graph

n = Number of persons, $p_1 \dots p_n$.

Person p_i has:

- Gender = {*male, female, unknown*}
- Relatives:

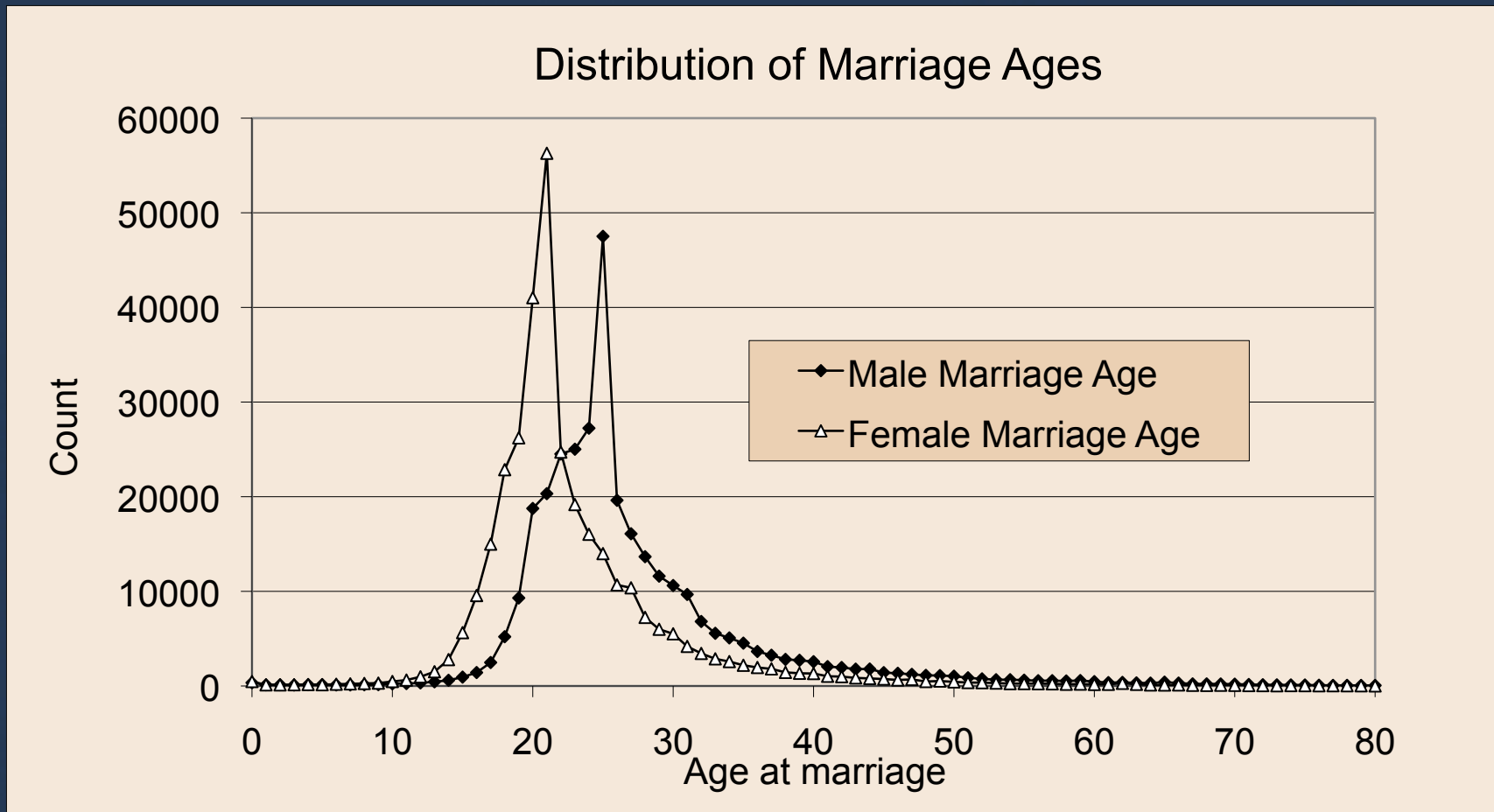


Deriving *Deltas*

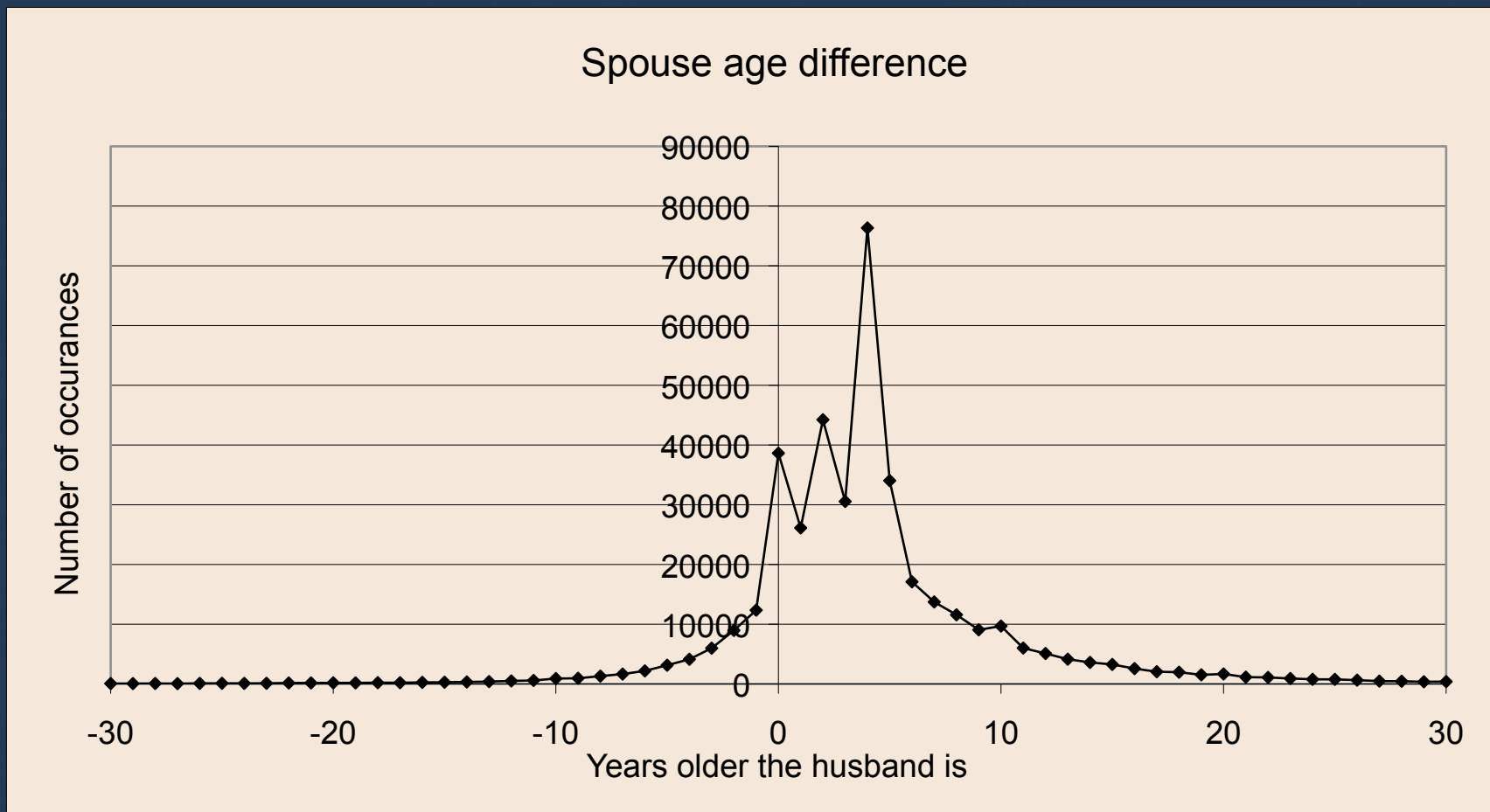
5-D array of cases from 15M people

1. **Target event:** *birth, marriage, death (single), death (married)*
2. **Relative type:** *individual, father, mother, spouse, child.*
3. **Source event:** *birth, christening, marriage, death/burial, other.*
4. **Gender:** *male, female, either/unknown*
5. **Exactness:** *specific (3 Jan 1820), year-only.*

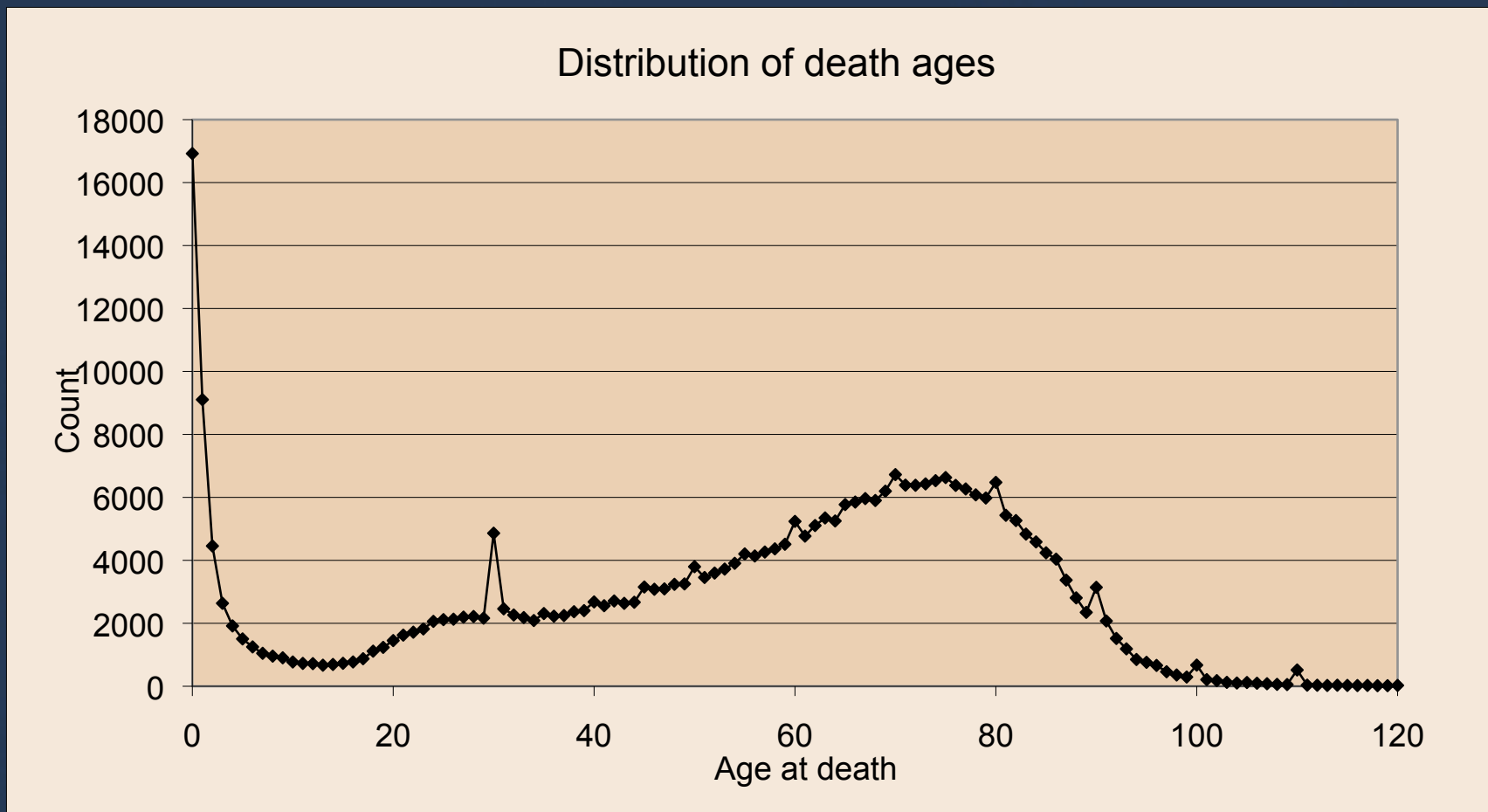
delta(birth, ind, marriage, {m,f}, exact)



Delta(birth, spouse, birth, male, exact)



Delta(death, ind, birth, male, exact)



Drop outliers

Drop top and bottom 1% => 98%

*delta(birth, individual, marriage,
male, specific)=17..63*

*delta(birth, individual, marriage,
female, specific)=14..52*

Delta tables

myEvent	Relative	Event	Year-only date						Specific (day, month, year) date					
			Male		Female		Either		Male		Female		Either	
Birth	Ind	Birth	0	0	0	0	0	0	0	0	0	0	0	0
Birth	Ind	Chr	-6	44	-7	36	-6	42	0	5	0	4	0	5
Birth	Ind	Death	0	97	0	100	0	99	0	92	0	94	0	93
Birth	Ind	Marriage	12	64	8	55	10	60	17	63	14	52	15	60
Birth	Father	Birth	-62	-14	-61	-14	-61	-14	-56	-19	-56	-19	-56	-19
Birth	Father	Chr	-63	21	-62	25	-63	22	-61	-1	-60	-3	-60	-1
Birth	Father	Death	-6	70	-5	69	-5	70	-1	65	-1	65	-1	65
Birth	Father	Marriage	-30	17	-29	16	-30	16	-25	5	-25	5	-25	5
Birth	Mother	Birth	-52	-10	-51	-10	-51	-10	-45	-17	-45	-17	-45	-17
Birth	Mother	Chr	-51	32	-53	32	-52	32	-46	0	-47	-1	-47	-1
Birth	Mother	Death	-5	75	-5	75	-5	75	0	70	0	70	0	70
Birth	Spouse	Birth	-12	29	-29	12	-24	24	-10	24	-24	10	-21	21
Birth	Spouse	Chr	-13	38	-31	31	-27	35	-14	39	-27	29	-25	35
Birth	Spouse	Death	21	110	18	100	19	107	25	106	23	93	24	103
Birth	Child	Birth	10	60	7	51	8	57	18	55	16	44	16	52
Birth	Child	Chr	17	63	14	54	15	59	18	61	16	52	16	59
Birth	Child	Death	24	139	21	133	22	136	23	136	20	129	21	133
Birth	Child	Marriage	28	101	25	95	26	98	37	94	34	88	35	92

Table 1. Deltas for calculating estimated birth date ranges from relatives' events.

Delta tables

			Year-only date				Specific (day, month, year) date							
myEvent	Relative	Event	Male		Female	Either		Male		Female	Either			
Marriage	Ind	Birth	-64	-12	-55	-8	-60	-10	-63	-17	-52	-14	-60	-15
Marriage	Ind	Chr	-69	19	-56	16	-65	18	-66	0	-54	0	-62	0
Marriage	Ind	Death	-2	78	-1	81	-1	80	0	70	0	75	0	73
Marriage	Ind	Marriage	0	0	0	0	0	0	0	0	0	0	0	0
Marriage	Father	Birth	-106	-34	-96	-30	-102	-31	-100	-40	-91	-37	-97	-38
Marriage	Father	Chr	-107	-6	-99	2	-106	-1	-107	-24	-97	-22	-104	-23
Marriage	Father	Death	-50	46	-42	50	-47	48	-44	43	-34	46	-41	44
Marriage	Father	Marriage	-78	-2	-68	0	-75	-1	-74	-14	-64	-11	-70	-12
Marriage	Mother	Birth	-100	-30	-90	-26	-96	-28	-94	-37	-84	-34	-90	-35
Marriage	Mother	Chr	-99	-2	-92	3	-97	1	-101	-18	-89	-12	-97	-15
Marriage	Mother	Death	-51	51	-43	54	-48	53	-42	48	-33	51	-38	50
Marriage	Spouse	Birth	-61	1	-66	-9	-64	-2	-60	-6	-65	-15	-63	-9
Marriage	Spouse	Chr	-64	19	-72	19	-69	19	-62	4	-68	2	-66	3
Marriage	Spouse	Death	-14	83	-11	78	-13	80	-10	78	-7	71	-9	76
Marriage	Child	Birth	-34	31	-27	27	-31	29	-31	26	-22	23	-28	25
Marriage	Child	Chr	-33	47	-25	45	-30	46	-31	31	-22	27	-27	29
Marriage	Child	Death	-1	111	1	110	0	110	-7	109	0	107	-3	108
Marriage	Child	Marriage	-15	74	-8	73	-12	74	-12	69	-5	67	-9	68

Table 2. Deltas for calculating estimated marriage date ranges from relatives' events.

			Year-only date						Specific (day, month, year) date					
myEvent	Relative	Event	Male		Female		Either		Male		Female		Either	
Death(married)	Ind.	Birth	-100	-21	-101	-18	-100	-19	-95	-25	-97	-20	-96	-22
Death(married)	Ind.	Chr.	-93	-10	-92	-14	-93	-13	-94	-2	-95	-2	-95	-3
Death(married)	Ind.	Death	0	0	0	0	0	0	0	0	0	0	0	0
Death(married)	Ind.	Marriage	-78	2	-81	1	-80	1	-70	0	-75	0	-73	0
Death(married)	Father	Birth	-140	-47	-140	-44	-140	-45	-138	-52	-141	-47	-140	-50
Death(married)	Father	Chr.	-136	-36	-134	-32	-135	-35	-140	-38	-141	-36	-141	-37
Death(married)	Father	Death	-83	25	-83	29	-83	27	-80	24	-82	29	-81	26
Death(married)	Father	Marriage	-111	-23	-112	-18	-111	-20	-109	-26	-112	-22	-110	-23
Death(married)	Mother	Birth	-133	-43	-133	-40	-133	-41	-130	-49	-133	-44	-131	-46
Death(married)	Mother	Chr.	-127	-38	-127	-35	-127	-37	-132	-32	-133	-30	-133	-31
Death(married)	Mother	Death	-84	29	-84	34	-84	31	-80	28	-82	33	-81	30
Death(married)	Spouse	Birth	-100	-18	-110	-21	-107	-19	-93	-23	-106	-25	-103	-24
Death(married)	Spouse	Chr.	-88	-15	-99	-19	-96	-16	-93	-16	-105	-18	-102	-17
Death(married)	Spouse	Death	-53	56	-56	53	-55	55	-51	53	-53	51	-52	52
Death(married)	Child	Birth	-76	6	-80	4	-78	5	-67	0	-72	0	-70	0
Death(married)	Child	Chr.	-67	17	-70	12	-68	15	-65	6	-67	3	-66	5
Death(married)	Child	Death	-52	84	-58	84	-54	84	-54	80	-59	80	-56	80
Death(married)	Child	Marriage	-51	50	-55	50	-53	50	-46	40	-51	39	-49	39
Death(single)	Ind.	Birth	-99	0	-100	0	-100	0	-90	0	-93	0	-91	0
Death(single)	Ind.	Chr.	-96	0	-97	0	-96	0	-83	0	-84	0	-83	0
Death(single)	Ind.	Death	0	0	0	0	0	0	0	0	0	0	0	0
Death(single)	Father	Birth	-139	-22	-140	-22	-139	-22	-134	-22	-137	-23	-135	-22
Death(single)	Father	Chr.	-140	-3	-144	-9	-142	-6	-132	-21	-134	-20	-133	-21
Death(single)	Father	Death	-79	59	-81	60	-79	60	-73	59	-76	59	-74	59
Death(single)	Father	Marriage	-110	0	-112	0	-111	0	-103	0	-106	0	-105	0
Death(single)	Mother	Birth	-132	-19	-134	-19	-133	-19	-125	-19	-128	-19	-127	-19
Death(single)	Mother	Chr.	-133	-3	-139	-6	-135	-4	-125	-17	-127	-17	-126	-17
Death(single)	Mother	Death	-79	64	-82	65	-80	65	-73	64	-75	64	-73	64

Table 3. Deltas for calculating estimated death date ranges from relatives' events.

Calculating ranges

From year and delta:

range.min = eventYear - delta.max

range.max = eventYear - delta.min

Father death= 1800

Delta = -1..65

=> Birth = 1800-65..1800-(-1)

= 1735..1801

Calculating ranges

From *range* and delta:

range.min = *eventRange.min* - *delta.max*

range.max = *eventRange.max* - *delta.min*

Father death = 1800..1820

Delta = -1..65

=> Birth = 1800-65..1820-(-1)

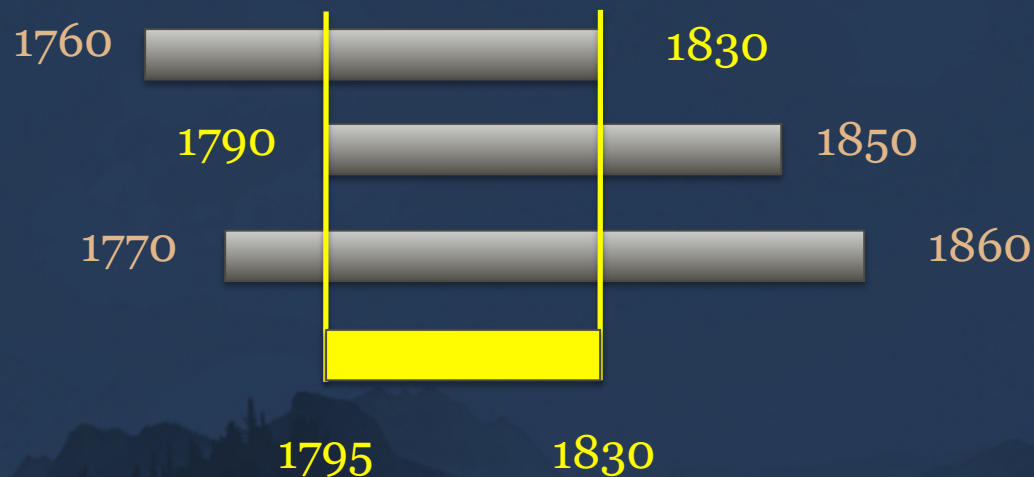
= 1735..1821

Combining evidence: Intersecting Ranges

From *range* and delta:

$range.min = \max(range_i.min)$

$range.max = \min(range_i.max)$



Ignoring Conflicting Data: Voting



Uses of date propagation

- Person matching
- “Reasonable” search results
- Living calculation
 - Propagate ranges to everyone.
 - Latest death year from own death year or (latest birth year + 110)
 - (Latest death year) < now => dead.

Evaluating Living Calculation

- Prune graph at 1900
 - Remove events after 1900
 - Remember death events
 - Remove people, spouses, descendants born after 1900
- Do date propagation
- Compare “estimated living” vs. known death dates

Evaluating Living Calculation

- Estimated and actual death year both before 1900 => "correct dead"
- Estimated and actual death year both after 1900 => "correct living"
- Estimated death year < 1900, actual > 1900 => "false dead" / "leaked living data"
- Estimated death year > 1900, actual < 1900 => "false living" / "(unnecessarily) hidden data"

Empirical Results

	Year Propagation		Range Propagation	
	count	percent	count	percent
Correct dead	1492	26.69%	1517	27.14%
Correct living	3458	61.86%	3194	57.14%
False dead (“Leaked living”)	36	0.64%	11	0.20%
False living (“Hidden dead”)	604	10.81%	868	15.53%

Future Research

- Larger sample size.
- Propagate *probability distributions*
 - ± 100 -year counts per range.
 - Use convolution
 - Renormalize after each iteration
 - Trim ranges at the end.




FAMILYSEARCH
WHERE GENERATIONS MEET

Thank You.

Randy Wilson

wilsonr@familysearch.org

Sponsored by:



NEW ENGLAND HISTORIC GENEALOGICAL SOCIETY

