Family Reconstitution and Nominal Linkage of parish registers in a Relational Database

Gill Newton

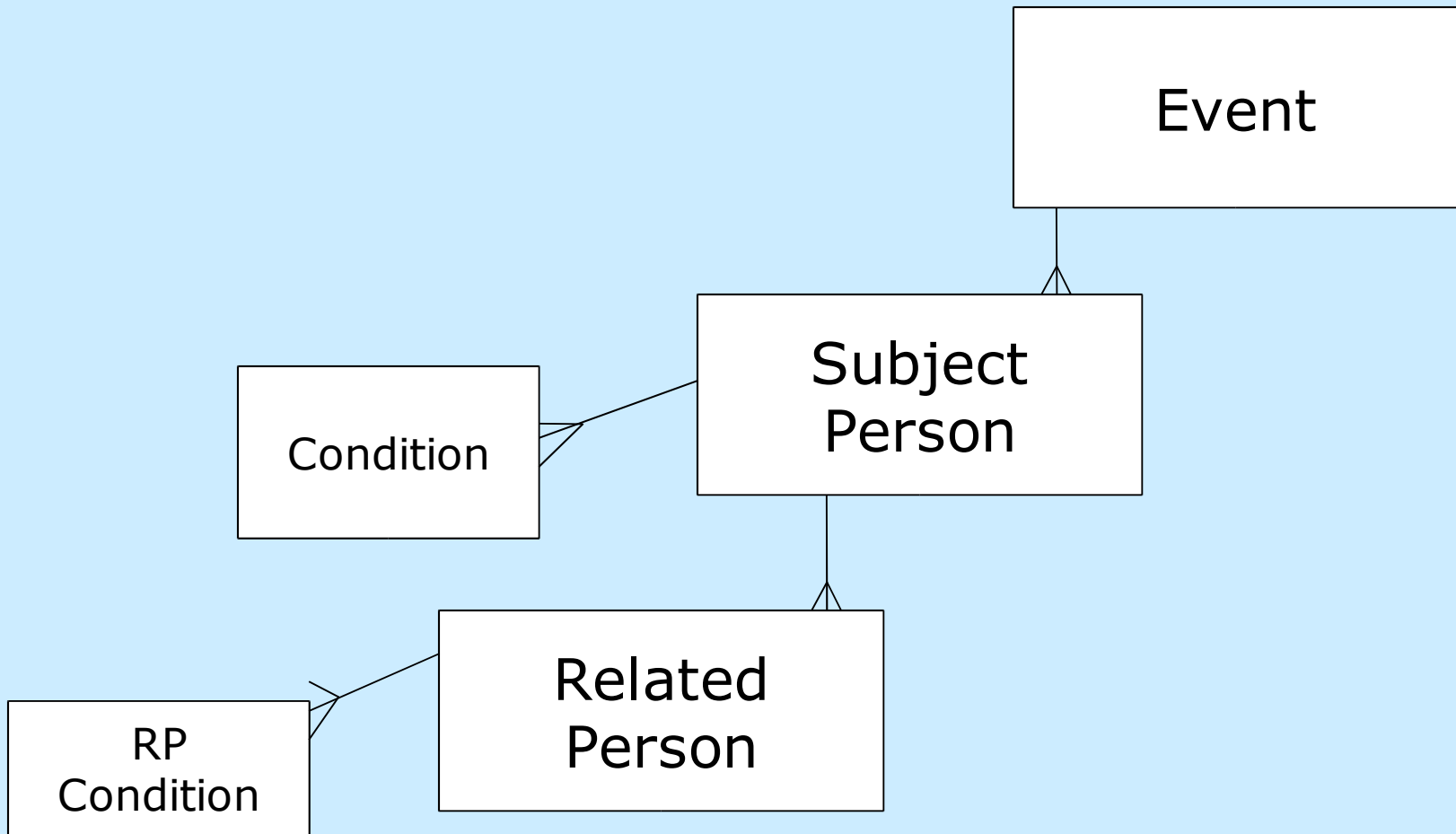Cambridge Group for the History of Population and Social Structure

Modelling the Parish Registers

- no *pro forma* layout in source

- PEFs for manual family reconstitution give framework for what data to expect
  BUT...
  - source registers problematic
  - format not optimal for database
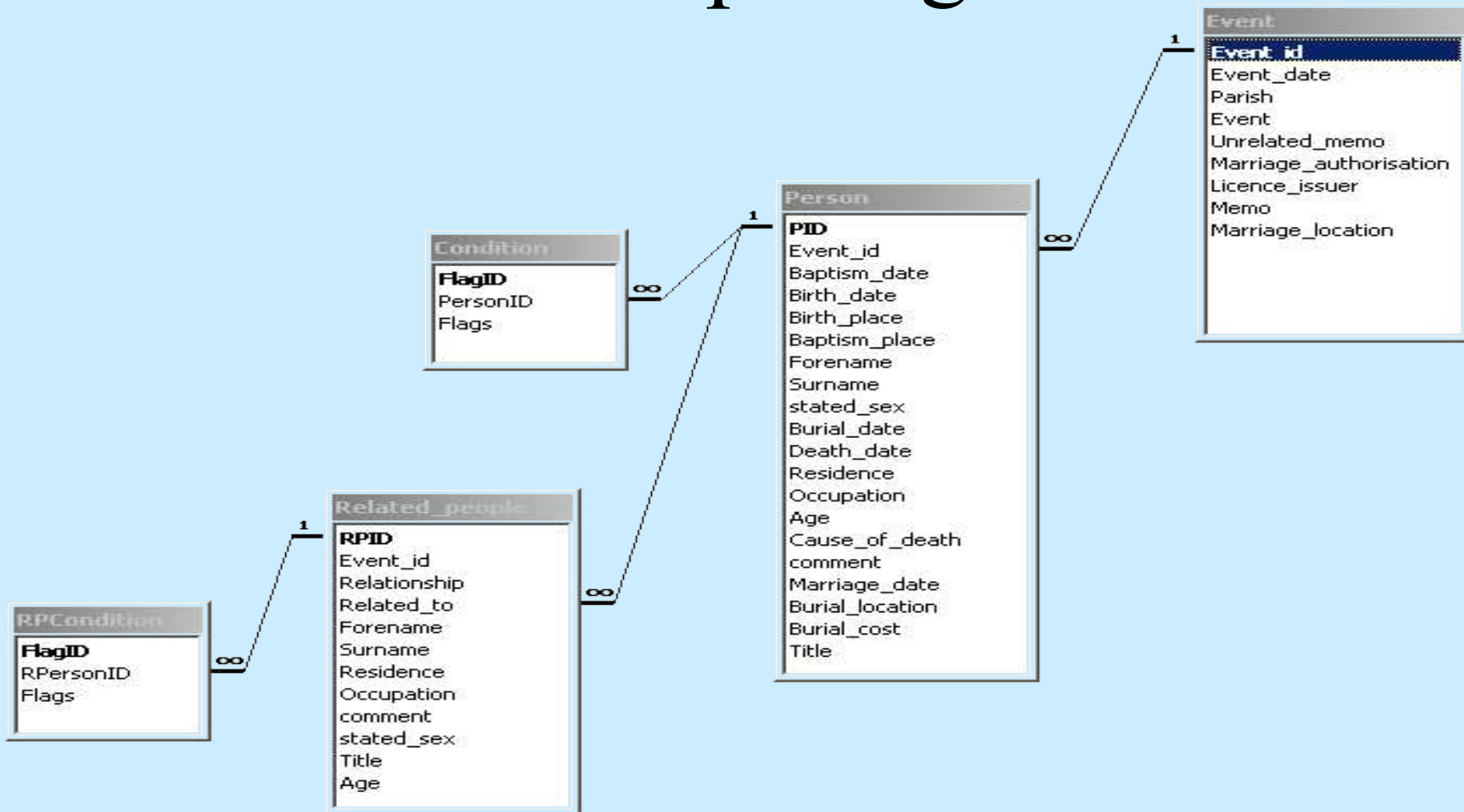 - project focus not only family reconstitution

The model

- organised around parish register events
  - Baptism, Marriage, Burial separate for input, as in manual FR

- aim to capture all data in the registers

- people are the base unit
  - before the FR cannot be family based
  - to allow representation of non-familial relationships (eg employer-employee)
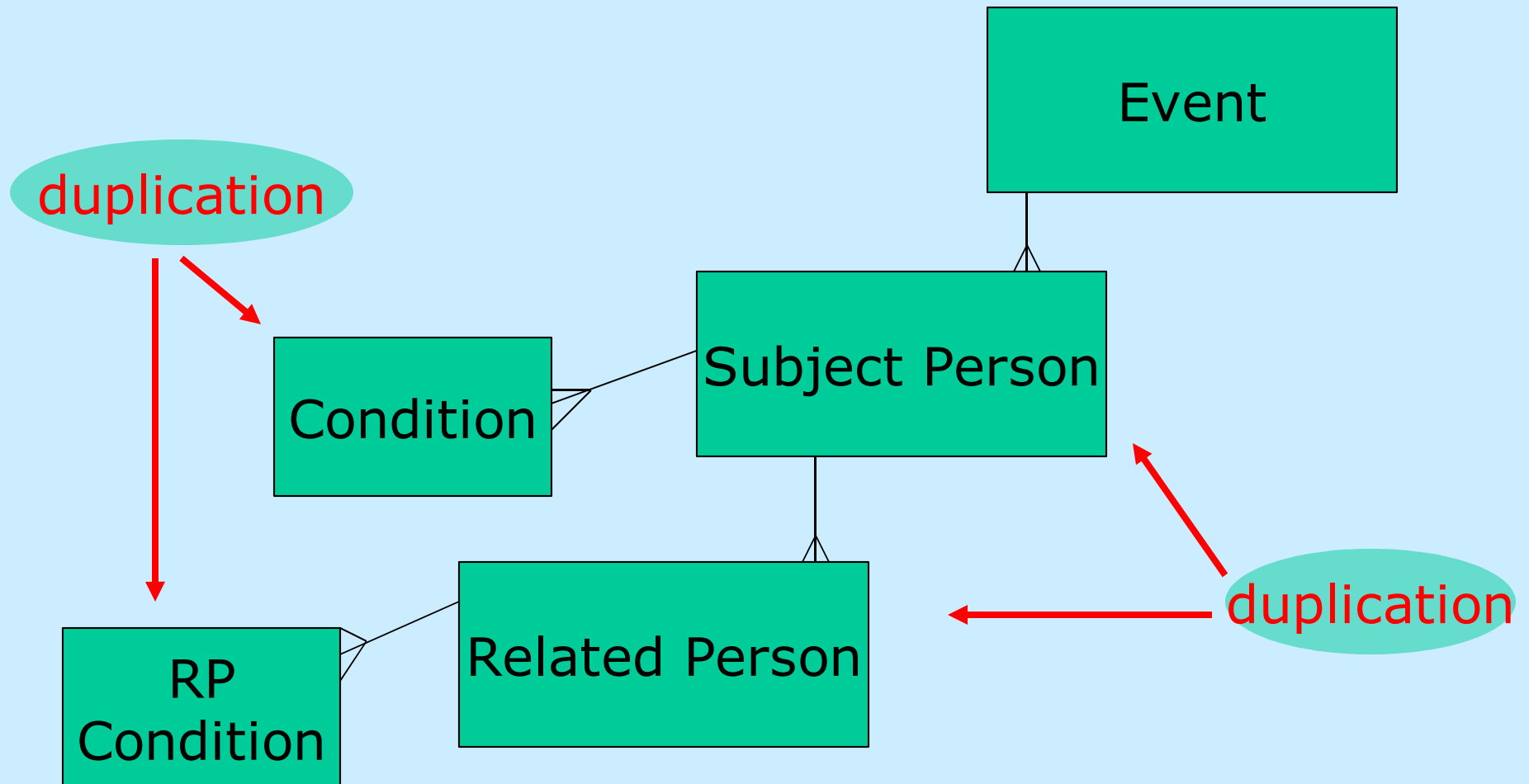
From parish registers to database: Entity Relationship diagram

Actual database Entity Relationship Diagram

Event
- **Event_id**
- Event_date
- Parish
- Event
- Unrelated_memo
- Marriage_authorisation
- Licence_issuer
- Memo
- Marriage_location

Condition
- **FlagID**
- PersonID
- Flags

Person
- **PID**
- Event_id
- Baptism_date
- Birth_date
- Birth_place
- Baptism_place
- Forename
- Surname
- stated_sex
- Burial_date
- Death_date
- Residence
- Occupation
- Age
- Cause_of_death
- comment
- Marriage_date
- Burial_location
- Burial_cost
- Title

Related_people
- **RPID**
- Event_id
- Relationship
- Related_to
- Forename
- Surname
- Residence
- Occupation
- comment
- stated_sex
- Title
- Age

RPCondition
- **FlagID**
- RPersonID
- Flags

Normalising the parish register model

Data capture: Forms for speed and accuracy

- designed to minimise keystrokes, maximise accuracy
  - sequence of fields mirrors "usual" record structure of source and is flexible
  - programmed to allow tab movement with minimal mouse usage
  - constraints on input eg no null relationships permitted

- time saved allows us to standardise AND keep the original orthography
  - freetext fields for original orthography (eg Occupation)
  - fields constrained to lookup tables (eg Condition)
  - deliberate duplication between the two (eg servants)

Extract from matrix showing lookup field:freetext field dependencies

person/related person lookups	lookup field					freetext field			
	Stated sex	Title	Condition	Relationship	Residence	Occupation	Comments	Residence	Age
Alderman	x	x	x			x			
alias			x				x		
ancient/old/aged			x						x
bachelor			x						
apprentice (of)			x	x		x			
Baron	x	x	x			*x			
baron of the Exchequer			x			x			
baronet			x			x			
black			x				x		
boy	x								
brother (of)	x			x					
Captain	x	x	*x?			*x			
child (of)			x	x					
christened			x						
churchwarden			x			x			
citizen of London			x			x			
clerk			x			x			
Colonel	x	x	*x?			*x			
common councilman			x			x			
Dame	x	x	*x?			*x			
daughter (of)	x			x					
deceased			x						
died presently			x						
doctor (Dr)	x	x	x			x			
Duke	x	x	*x?			*x			
Duchess	x	x	x			*x			
esquire			x			x			
father (of)	x	x	*x?	x		*x			
female	x								
foreigner			x					x	
foundling			x						
free of the			x			x			
generosus			x			x			

Event ID	1555	Memo		▼	Unrelated memo	
Event date	23/08/1625					
Parish	All Hallows Hone ▼					
Event	Burial ▼					

Burial Subject

PID	Burial date	Death date	Title	Forename	Surname	Stated sex	Occupation	Condition
1714	August 1625		Mrs ▼	Marie	Holder	Mrs ▼		widow(er) ▼
								▼

Residence	Cause of death	Age	Burial location	Burial cost	Comment
			in the churchyard against Milk Streete Church ▼	vis iiijd	

Record: ◄◄ ◄ 1 ► ►► ►* of 1

Related People

RPID	Related to PID	Relationship	Title	Forename	Surname	Occupation	Condition
(AutoNumber)	1714	▼	▼				▼

Stated sex	Residence	Age	Comment
unstated ▼	▼		

Record: ◄◄ ◄ 1 ► ►► ►* of 1

New Record   Delete Event

Linking people: data points

- Names
  - not only of subject but also parents/spouse names
  - titles
  - match by soundmatching algorithm/by lookup table
- Dates
  - permissible intervals between demographic events subject to rules
  - almost no age information is directly given in the registers
- Residence
  - of/not of this parish (marriage records)
  - stranger (burial/baptism records)
- Occupation
  - Some indicator of age (burial records)
  - gross mismatches disallowed

Tackling varying orthography

In favour of standardising...
- aids searching (querying) the data
    - allows earlier analysis of the data
- aids linking the data
- speeds up data entry

In favour of preserving...
- gives solid foundation for uncertain data
    - dangers of assumptions/interpretation
- philological interest
- to ease checking

forename	DMET1	DMET2	sex_in_data	working_se	name_expanded	workingDMET
Ellice	ALS		M	m		ALS
Elizeus	ALSS		m	m	Elias	ALS
Alce	ALS		f	f		ALS
Elias	ALS		m	m		ALS
Elyas	ALS		m	m		ALS
Elis	ALS		M	m		ALS
Ellis	ALS		m	m		ALS
Alys	ALS		f	f		ALS
Alxe	ALKS		F	x	Alice	ALS
Elisha	ALX		m	m	Elias	ALS
Ales	ALS		f	f		ALS
Alic	ALK		f	f	Alice	ALS
Alies	ALS		f	f		ALS
Alice	ALS		f	f		ALS
Alse	ALS		f	f		ALS
Alis	ALS		f	f		ALS
Alles	ALS		f	f		ALS
Allice	ALS		f	f		ALS
Allyce	ALS		F	f		ALS
Allys	ALS		F	f		ALS
Eliabethe	ALP0	ALPT	f	f	Elizabeth	ALSP
Eleazabeth	ALSP		f	f		ALSP
Elezabeth	ALSP		f	f		ALSP
Elezabethe	ALSP		F	f		ALSP
Elezebeth	ALSP		f	f		ALSP
Elisabethe	ALSP		f	f		ALSP
Eleizabeth	ALSP		f	f		ALSP
Elisabeth	ALSP		f	f		ALSP
Elixabeth	ALKS		F	f	Elizabeth	ALSP
Eliza.	ALS		F	f	Elizabeth	ALSP

Cheapside parishes events per year: Aggregative totals

BMB 9 Year Moving Averages
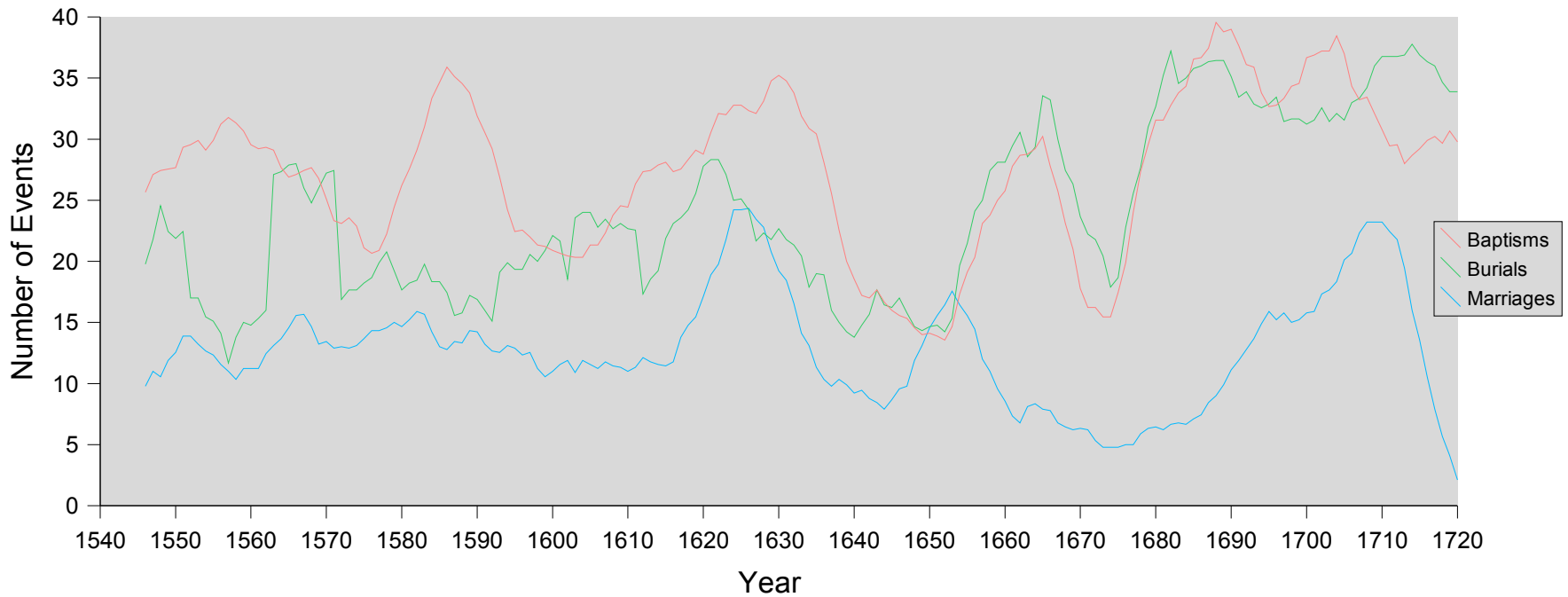
NB: St Mary le Bowe registers 1630-50 (bap and bur) and 1630-1670 (mar) missing

Quality of Cheapside data 1540-1720: Baptisms

- 86% of St Mary le Bowe baptisms have a named parent
- 97% of All Hallows and 46% of St Pancras baptisms have a named parent
- After 1675 (c1640 for St Pancras), mother's name routinely given

BUT...

- St Mary le Bowe baptisms 1631-53 missing
- All Hallows 1651-65 and 1671-82 missing

Quality of Cheapside data 1540-1720: Marriages

- 53% of St Mary le Bowe marriages give residence information

- 75% of All Hallows and 46% of St Pancras marriages give residence information

- Combined parishes marriages 1675-1720 give residence information, some bride's fathers and marital status if widow(er).

BUT...

- St Mary le Bowe marriages 1631-75 missing
- All Hallows 1666-75 marriages missing

Quality of Cheapside data 1540-1720: Burials

- Could be crux of the problem
- ...But quality is difficult to evaluate
  - Distinguishing between "short form" (names only) and male adults (including heads)
- St Mary le Bowe no register 1631-53 and:
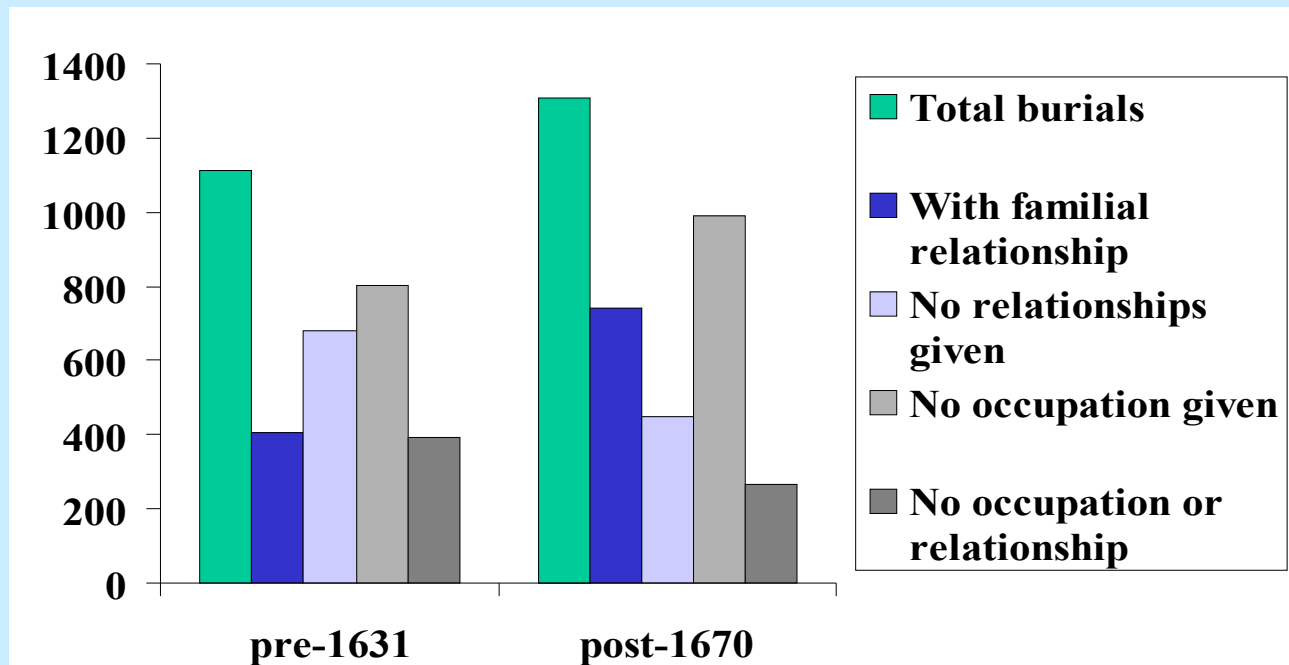
Family Reconstitution:
the method in theory

- 10 steps in 'Nominal Record Linkage by Computer' (Wrigley, E A and Schofield, R S, *Idenfifying People in the Past*, London, 1973)
    1. BAP-MAR [parents]
    2. MAR-BUR [surviving spouse]
    3. BAP-MAR [parents] revision
    4. BAP-MAR [marriage partners to their own baptisms]
    5. BAP/MAR-BUR
    6. BAP/MAR-BUR revision
    7. MAR-MAR [male]
    8. MAR-MAR [male] revision
    9. MAR-MAR [female]
    10. MAR-MAR [female] revision
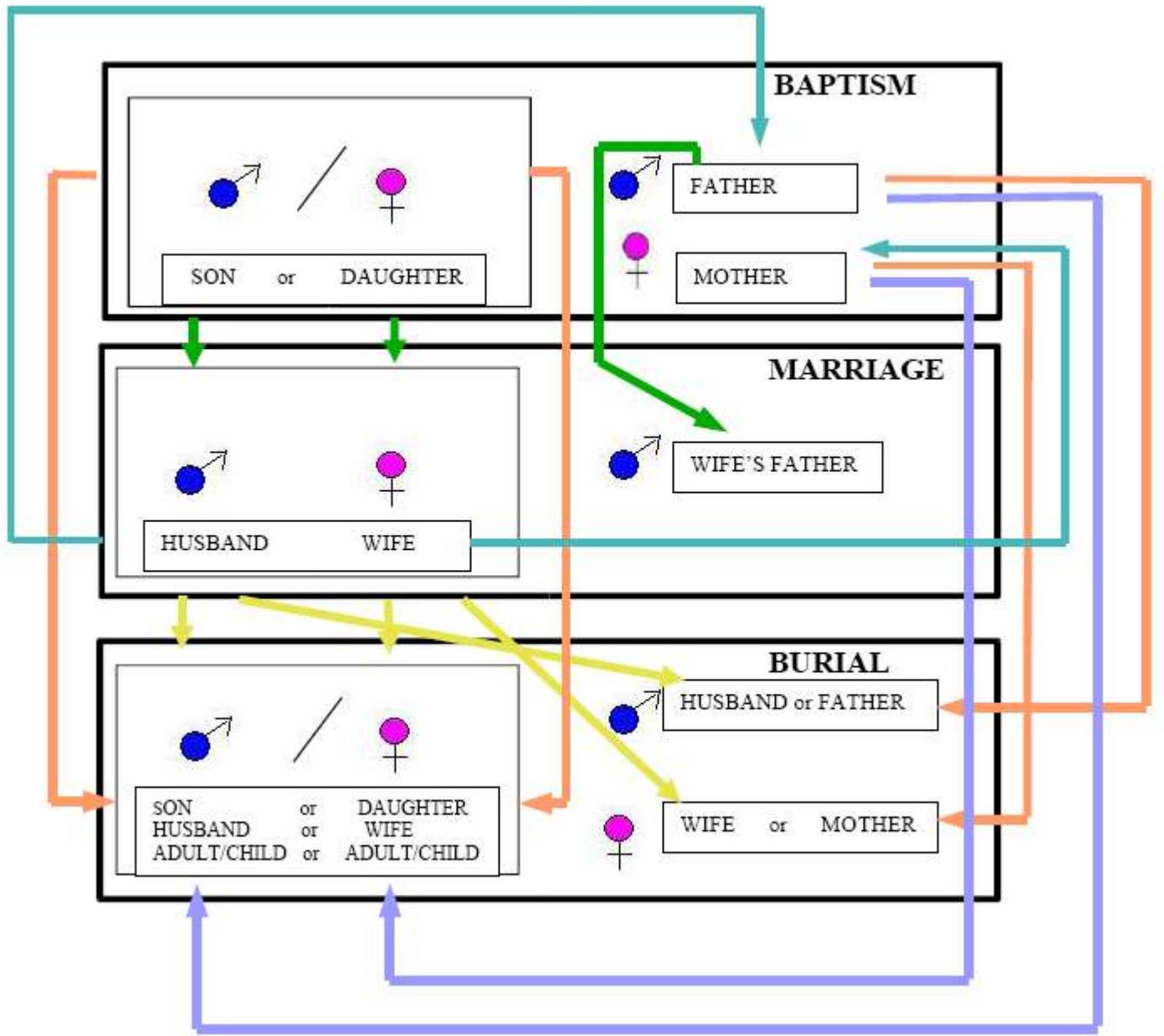
Family Reconstitution of Cheapside: Diagram

Decision making: matchscoring

- Where multiple possible matches exist, must choose
- BUT seldom have the choice of more than one match
- Difficult to base weighting on sample of correctly linked data
  - small sample size
  - what is correct?
- ...so positive/negative weighting by data point

- sometimes necessary to carry out matchscoring by hand eg
  - eg closeness of name matching

Cheapside reconsitution:
MAR-BAP linkage query example - Resulting table, with matchscores

baprevised	mardate	FamiliesW	Marflatfile	burialdate	ageatm	forenam	surname	Hforenam	Hsurname	sex		Hres	Hconditi	condition1	matchscore
15/06/1544	04/06/1564	St Mary le E	All Hallows		20	John	Bartelmew	John	Bartyllmew	son					0
23/02/1551	18/06/1567	St Mary le E	St Mary le E		16	John	Ferar	John	Farrer	unstated					1
18/01/1551	02/01/1570	St Mary le E	All Hallows		19	William	Woreley	Wyllyam	Earely	unstated					-5
17/06/1547	08/10/1572	St Mary le E	St Mary le E		25	William	Bonner	Wm	Bonner	unstated					1
20/02/1543	03/01/1574	St Mary le E	St Mary le E		31	Thomas	Farrer	Thomas	Freere	son					-5
20/04/1544	18/01/1574	All Hallows	All Hallows		30	Nicolas	Fuller	Nicholas	Fullor	son		of Gray's Inn			1
05/09/1555	25/04/1574	St Pancras	St Pancras	03/01/1563	19	Richard	Cocke[\]	Rychard	Cooke	son					1
12/09/1543	26/02/1576	All Hallows	All Hallows		33	Thomas	Goter	Thomas	Cawdray	son		par. of St Grygores			-5
04/09/1542	16/07/1576	All Hallows	St Mary le E		34	Cutberd	Fuller	Cutberd	Fulwer	son				twins/multiple	0
08/08/1554	01/12/1577	All Hallows	St Mary le E		23	Thomas	Blunt	Thomas	Bowland	son					-5
11/10/1539	01/03/1579	All Hallows	St Pancras		40	Thomas	Smythe	Thomas	Smyth	son					0
01/11/1547	26/01/1580	St Mary le E	St Mary le E		33	William	Lee	William	Lye	unstated					0
07/05/1540	19/06/1580	St Pancras	St Pancras		40	Thomas	Merbury	Thomas	Marberye	son					1
11/10/1561	07/07/1582	St Mary le E	St Pancras		21	Thomas	Allin	Thomas	Allen	unstated					0
05/11/1558	16/12/1582	All Hallows	St Pancras		24	Rowland	Mosly	Rouland	Mosley	son					0
28/12/1550	29/07/1585	St Mary le E	St Pancras		35	William	Walker	William	Walker	unstated		this parish			0
19/03/1563	24/04/1587	All Hallows	St Pancras		24	Edward	Barnes	Edwarde	Barnes	son					0
07/11/1559	01/02/1588	St Pancras	St Pancras		29	Edward	Peersone	Edward	Pierson	son					1
19/04/1564	29/09/1588	St Mary le E	All Hallows		24	Wm	Gibson	Willyam	Gibson	son		S. Alphage			-5
31/08/1553	28/11/1592	St Mary le E	St Mary le E		39	John	Cox	John	Cockes	unstated					1

Cheapside reconstitution in practice: the mechanics

- series of queries implemented to arrive at listing of families
    - forms secondary analysis layer in database
    - each query kept relatively simple
- all queries saved and grouped by stage of linkage
- queries have demographic constraints embedded: self-documenting
- ...yet only partly automated
    - each query stands in isolation: takes human input to run the chains of queries
    - matchscoring hopefully keeps decision-making objective, but the need to make a decision is human-directed
    - difficulty of rolling through updates on primary data
- two further parishes on which to improve the process

Cheapside reconstitution: evaluation

- as close to theory as practicable but with emphasis on reconstructing relationships even where whole families are not in observation
- primary purpose not of demographic analysis but as a framework onto which to link other sources: building histories, tax, wills
- "community reconstruction" is desired effect
- use of biological best fit constraints rather than absolutes to supplement sparse data
  - eg 25 year best fit maximum fertility period for marriage
    - little age information available for mothers
    - less than 35 to prevent intergenerational matches
  - this has implications for statistical analysis