DEMOGRAPHIC RESEARCH

Research Article

A general temporal data model and the structured population event history register

Samuel J. Clark

Table of Contents

A general temporal data model and the structured population event history register

Samuel J. Clark[1]

Abstract

At this time there are 37 demographic surveillance system sites active in sub-Saharan Africa, Asia and Central America, and this number is growing continuously. These sites and other longitudinal population and health research projects generate large quantities of complex temporal data in order to describe, explain and investigate the event histories of individuals and the populations they constitute. This article presents possible solutions to some of the key data management challenges associated with those data. The fundamental components of a temporal system are identified and both they and their relationships to each other are given simple, standardized definitions. Further, a metadata framework is proposed to endow this abstract generalization with specific meaning and to bind the definitions of the data to the data themselves. The result is a temporal data model that is generalized, conceptually tractable, and inherently contains a full description of the primary data it organizes. Individual databases utilizing this temporal data model can be customized to suit the needs of their operators without modifying the underlying design of the database or sacrificing the potential to transparently share compatible subsets of their data with other similar databases. A practical working relational database design based on this general temporal data model is presented and demonstrated. This work has arisen out of experience with demographic surveillance in the developing world, and although the challenges and their solutions are more general, the discussion is organized around applications in demographic surveillance. An appendix contains detailed examples and working prototype databases that implement the examples discussed in the text.

---

[1] Department of Sociology, University of Washington; Institute of Behavioral Science (IBS), University of Colorado at Boulder; MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, University of the Witwatersrand.
Corresponding Addresses: work@samclark.net.

1. Introduction

The questions that interest population and health scientists are becoming increasingly complex and require longer periods of more intense observation to explore. Investigations of this type generate significant quantities of complex temporal data that describe the interrelated histories of people and groups of people. A case in point are the 37 mostly new demographic surveillance system (DSS) sites that are members of the INDEPTH Network (INDEPTH Network, 2004) and continuously generate data describing roughly one million people in eighteen different countries, as well as the growing number of prevention, intervention and drug trials (For example: HPTN, 2006; HVTN, 2006; IAVI, 2006; SAAVI, 2006). Trials of this sort – vaccine, behavioral modification, mosquito-control, poverty alleviation, micronutrient supplementation, etc. – require the long-term study of well-defined populations so that events can be correctly sequenced and the effects of interventions can be properly identified and disentangled from possible confounding factors.

Effective management and analysis of these data are essential to the success of these studies. Poor data management results in corrupt data that are difficult to access and analyze, and this reduces the overall productivity and validity of studies and makes it difficult to share or compare data with other similar studies.

At the core of all longitudinal data management systems is a *temporal database* that is able to store and manipulate the data collected by a longitudinal project - and it is often the case that this database is based on an idiosyncratic design that has evolved in an *ad hoc* fashion over many years. For individual projects this results in poorly functioning databases that allow complex inaccuracies and errors to accumulate in the data, and for the group of longitudinal projects as a whole, the result is a collection of largely incompatible longitudinal databases whose data cannot be easily shared and analyzed together. This limits the usefulness of data collected by individual projects and largely denies the potential synergy that sharing, comparing and pooling data from multiple longitudinal studies could afford.

One solution is to develop: 1) standard definitions for temporal data and 2) a standard temporal design for databases that store such information. Because there is such diversity in the studies conducted by longitudinal projects, both the standard data definition and database design must be general and flexible enough to define and manage a wide variety of data, and to be able to easily accommodate changes to the overall set of data collected and managed by an individual project as time progresses. Standards meeting these criteria will enable individual projects to manage their data in a conceptually consistent, accurate and well-documented way throughout their period of investigation. This in turn will lead to greater accuracy and productivity and provide the potential to easily share data with and among other sites that utilize the same

standards. This additional ability to easily share and pool data from multiple sites will increase the value of all the data and lower the barriers to designing and implementing prospective multi-site studies.

This article introduces some of the data management challenges created by complex temporal data and presents potential solutions. This work has arisen out of experience with demographic surveillance in the developing world, and although the challenges and their solutions are more general, the following discussion is organized around their application to demographic surveillance. Demographic surveillance involves recording the vital, nuptial and migration histories of well-defined populations over time[2].

1.1 The structure of temporal data

Longitudinal data are generated through many different study designs, including:

- linked, repeated cross-sectional surveys,
- panel studies,
- cohort studies,
- population laboratories (community-level population surveillance),
- vaccine, intervention, prevention and drug trials,
- environmental monitoring, and
- ecology laboratories.

Each of these have in common that they make repeated observations of the study subjects and link the data collected at each observation. Beyond this there is great variability in how the subjects are chosen, when and how they enter and exit the study, whether any information is collected that describes the interaction of study subjects with each other, and finally whether data are collected at more than one level of aggregation, perhaps at the household and community levels in addition to the individual level.

Temporal data have three main dimensions: the entity, its attributes, and time. The entity dimension relates to the set of items being described by the data. The attributes dimension relates to the set of attributes that describes each entity, some of which are constant while others vary with time. The time dimension relates to the temporal qualities of the data. For example, that entities come into existence, cease to exist,

---

[2] Typical DSS sites monitor geographically circumscribed populations of tens of thousands of people over periods of years or decades.

come under observation and are no longer observed. The time-evolving relationships between entities describe the structure and dynamics of the population and are difficult to model and manage in a general, efficient way.

In contrast, non-temporal data have only the first two dimensions, the entities and their attributes. These can be easily represented with the traditional two-dimensional table; each column associated with an attribute, and each row representing a different entity. Time adds a third dimension to the table so that the value of each attribute is recorded at all times. One can visualize this as a stack of two-dimensional tables with one two-dimensional layer for each instant in time. Attempting to implement this "stack of tables" solution would quickly reveal its significant limitations. Time is continuous and unbounded (see Section A.5) which would require an infinite number of layers in the stack, and values would be repeated many times for those attributes that remain constant over time. This would require an infinite amount of storage and would result in vast replication of data, with the potential for inconsistencies to arise when duplicate representations of the same fact are not identical. Additionally this model does not provide a tractable means to represent collections of entities (such as memberships of individuals in households or villages and their periods of residence at different locations), their relationships to one another as they change through time, or their relationships with entities of different types.

Clearly traditional two-dimensional tables cannot accurately and efficiently store the time dimension of data, and accordingly the representation and manipulation of temporal data has attracted a lot of attention.

1.2 Existing temporal frameworks

Over the past two decades computer scientists have expended much effort developing conceptual frameworks for temporal data (For example: Allen, 1983; Allen and Ferguson, 1994; Etzion et al., 1998; International Organization for Standardization, 2000; Jensen, 2000; Snodgrass, 2000; Spaccapietra et al., 1998). Few of these ideas have been incorporated into working database management systems that are widely available. Nevertheless, a standard terminology has been identified and is available as the "Consenus Glossary of Temporal Database Concepts" (Jensen et al., 1998), a basic set of standard temporal primitives and operators have been defined and largely incorporated into the current international standard for the Structured Query Language (SQL) (Gulutzan and Pelzer, 1999), and a recent volume is devoted entirely to the extension of the Relational Model of Data to incorporate temporal data (Date et al., 2002c). Meanwhile, population and health scientists have designed their own solutions using existing database tools. The majority of these have been developed on an *ad hoc,*

in-house basis by individual groups who needed an immediate solution to a specific data management need. Consequently it is not possible to easily share and compare the data they manage, or in most cases to even understand on what principles they operate because no published documentation is available.

A prominent exception is the Household Registration System (HRS) developed by Bruce MacLeod and colleagues in conjunction with the Navrongo DSS site in Ghana (MacLeod et al., 1996; Phillips et al., 2000). The HRS forms the basis of the data management system used by a number of the DSS sites, and as a result is the *de facto* standard data management system for DSS sites. Conceptually the HRS is built around the Reference Data Model (RDM) (Benzler et al., 1998), which is a temporal design for a relational database that can record the history of a human population. The RDM recognizes key events that determine transitions in a human life and key episodes that mark time intervals between events, for example a marriage or residence, during which a well-defined state is maintained. The RDM is able to record and manage social relationships, membership in social groups, residences at various locations, "status" observations, observation times, and all of the events necessary to define the population under observation and track its basic dynamics.

A drawback of the RDM and hence of the HRS is that it is an inflexible design[3]. It cannot be easily extended or modified without adding new database components or substantially modifying existing ones. As a result each site that uses the RDM in the form of the HRS has to invest time customizing the data management system to suit its own needs – albeit much less time and energy than if they developed their systems *de novo* instead of using the HRS as a starting point. This customization has resulted in a number of different implementations that are no longer compatible with one another; thus compromising one of the most important benefits of standardization – the ability to easily share and compare data stored in two or more systems based on the same standard.

2. Aim

This article presents a general temporal model of reality that unifies the representations of: 1) time, 2) the structure of a complex population of interconnected entities and 3)

---

[3] Another specific, important drawback of most HRS-based systems is that they use a combination of village number, household number and 'line number' (a personal ID defined with respect to a household) to identify unique individuals. This is convenient for field workers who can easily locate and identify individuals by "reading" their three-part ID number. Less convenient is that when an individual moves to another household or village their ID number changes. This leads to two separate and often unlinked representations of the same person, which in turn leads to duplication of information describing that person, and hence to the possibility of double-counting that person or their attributes during analysis.

descriptions of the individual entities that compose the population. The fundamental components of a temporal system are identified and both they and their relationships to each other are given simple, standardized definitions. A metadata[4] framework is proposed to endow this abstract generalization with specific meaning and to bind the definitions of the data to the data themselves.

Researchers can customize individual databases designed around the *General Temporal Data Model* (GTDM) without modifying the database schema[5] or its physical implementation[6], thus maintaining the potential to share compatible subsets of their data with other similar databases. The structure of the GTDM is highly normalized[7] meaning that individual facts are stored only once, thereby eliminating the possibility that duplicate representations of the same fact are different. In addition to documenting the primary data, the metadata make it possible to create general operations that affect the primary data automatically, customizing their effects based on the information contained in the metadata. This provides the potential to automate routine database management tasks. The GTDM naturally facilitates the definition of hierarchical groupings and is able to track the dynamics of those groups and their members, which is the information needed to analyze the social dynamics of human populations.

Although the GTDM is sufficiently general to be applied to any temporal system, we are interested specifically in its application to humans in the context of population and health studies. The remainder of this article will present:

- an explicitly temporal model of reality – the GTDM,
- a relational database implementation of the GTDM called the *Structured Population Event History Register* (SPEHR),
- a multi-site extension of SPEHR, and
- an appendix containing detailed working examples.

3. A general temporal model of reality

A temporal model is one that explicitly considers time as one of its components. The fundamental challenges emerge from the fact that time is universal, dense and unbounded, whereas most other entities in the model are discrete within both time and space and thereby only able to affect or be connected to a finite number of other

---

[4] Metadata – data that describe other data, see Section 4.1.
[5] Database schema – a description of the logical structure and organization of a database.
[6] Physical implementation – the manner in which information is organized and recorded on physical media such as hard disk drives.
[7] Normalized – a database design that minimizes replication of facts within the database, see Section 4.7.

entities. In contrast, time affects everything, everything experiences time, time has no bounds and time can be resolved with infinite precision.

The following sections present an example, describe the conceptual entities necessary to model a temporal reality and finally propose a general, integrated temporal model of reality. Section A.5 in the appendix contains detailed definitions of temporal elements that may be helpful while reading the following sections.

Figure 1: **Diagram of Event ⇔ Influence ⇔ State Example**
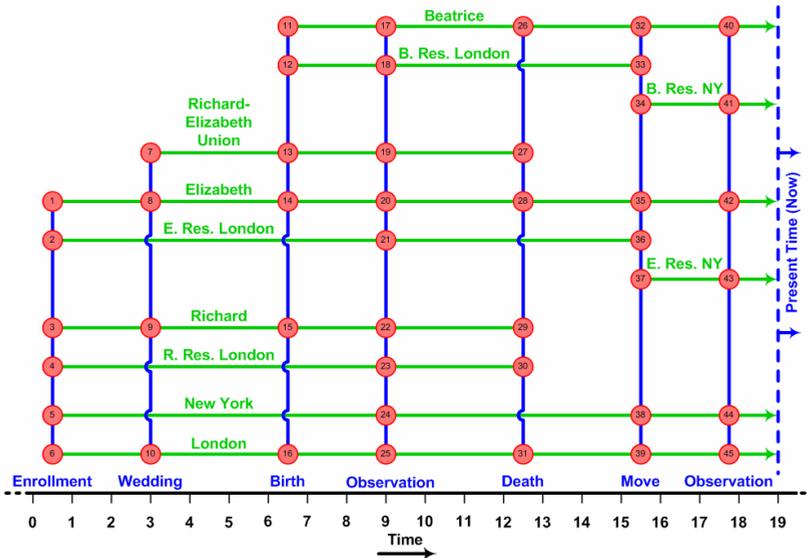
3.1 The New York – London example: Event ⇔ Influence ⇔ State

Imagine we are interested in recording the vital events, nuptial histories and migratory behavior of a small number of people living in New York and London, and assume that we have a universal clock to record the times when events occur. We initiate a small study at time 0.5 and enroll the two locations, New York and London; two people, Richard and Elizabeth; and because Richard and Elizabeth live in London, we initiate a residence for each at London. At time 3.0 a wedding occurs in London that joins Richard and Elizabeth as a couple and initiates their marital union. At time 6.5 in London Elizabeth gives birth to Beatrice, and at time 9.0 we visit our "sites" and make

an observation of all the existing entities enrolled in our study. At time 12.5 in London Richard dies and as a result of Richard's death Elizabeth and Beatrice move from London to New York at time 15.5. At time 17.8 we visit our "sites" again and make another observation of all the existing entities enrolled in our study, and the study continues to the present time, 19.0. Throughout the study we organize and record the information as events, influences and states.

Figure 1 displays a diagram of the information we collect. Time is recorded on the horizontal axis and is marked with equidistant positional markers from 0-19. Anticipating the discussion in the following three sections, horizontal (green) lines represent states, vertical (blue) lines represent events and the shaded (red) circles at the intersection of horizontal and vertical lines represent (potentially multiple) influences. States and events have descriptive labels while intersections between states and events where influences occur are numbered. Descriptions of the influences at each numbered intersection of a state and an event are contained in Table 1. The influence descriptions are written with reference to the states indicating how each state *is influenced* by each event. This example is implemented in a working database (see Section A.1) and is described in great detail in Section A.2.1 of the appendix.

3.2 Temporal entities: states

All entities that we wish to consider have a valid lifetime. This makes them states in the sense described below in A.5.1.2.4. They all have a well-defined beginning (start) and end (stop) with a *constant meaningful* state between those; most generally the state of "existing". Consequently we generalize the term *state* and use it as a concise name for temporal entities. States can represent both physical (e.g. people) and nonphysical (e.g. marital unions) entities.

The New York – London example contains four types of state that represent: 1) *people* who are *alive* from the time they are born until they die, e.g. Richard, Elizabeth and Beatrice; 2) *places* that are *meaningful locations* on the planet from the time they are enrolled in the study until the study ends, e.g. New York and London; 3) *unions* that are *sanctioned relationships* between two people from the time when they are licensed until they are terminated through annulment, divorce or death, e.g. Richard and Elizabeth's union; and 4) *residences* that are *durations of time when a person resides at a location* – e.g. Beatrice's residence at London. These states are listed as the rows of Table 1.

3.3 Temporal junctures: events

Events bring about the temporal change we wish to consider. As defined below in A.5.1.2.2 an event is a meaningful happening associated with a well-defined time point. Implicit in the meaning of an event is the change that it represents, and it is this notion of change that is most important to us. Events bring about or signify the beginning and ending of all states.

Events are thus temporal junctures: they form and dissolve all the relationships that exist between states, providing the means through which all states are joined. For example a birth affects at least: the infant, the mother, the father, the place where the birth takes place and the existing siblings. All of these states are linked to each other as a result of the *event of the birth*; some states were previously linked in different ways through the occurrence of a wedding and other births – but this new birth changes something for all of the states and creates a new set of links.

The New York – London example contains six types of event, briefly: 1) an *enrollment* event that initiates the states that describe the entities that are present at the beginning of the study, 2) a *wedding* that joins Richard and Elizabeth in a marriage and initiates their union state, 3) a *birth* that happens to Elizabeth and Richard and starts Beatrice's life, 4) a *death* that marks the end of Richard's life and residence state, 5) a *move* that takes Elizabeth and Beatrice from London to New York and terminates and initiates various residence states, and 6) two *observations*. These events are listed in the columns of Table 1.

3.4 Temporal nexus: influences

Influences form a temporal nexus between states and the events that affect them. Influences can be understood as the explicit representation of the links between states and events. In Figure 1 influences are represented by red circles at the junctures of states and events. Because it is possible for an event to influence more than one state, each event can be linked to many states. States are linked to each other through their individual links to the same event. Event histories of the states intersect when an event influences two or more states, and these intersections represent the formation and dissolution of relationships between the states.

Table 1: **Influences in London – New York, Event ⇔ Experience ⇔ State Example**

States	Enrollment	Wedding	Birth	Events: Observation @ 9	Death	Move	Observation @ 17
Beatrice			11 a. Person start: birth b. Person, residence start: birth	17 a. Person, observed: observation	26 a. Person, parent dies: death	32 a. Person, move away from: move (London) b. Person, move to: move (New York) c. Person, residence stop: move (London) d. Person, residence start: move (New York)	40 a. Person, observed: observation
Beatrice Residence at London			12 a. Residence start: birth	18 a. Residence, observed: observation		33 a. Residence stop: move	41 a. Residence, observed: observation
Beatrice Residence at New York						34 a. Residence start: move	
Richard-Elizabeth Union		7 a. Union start: wedding	13 a. Union, child born: birth	19 a. Union, observed: observation	27 a. Union stop: spouse dies: death		
Elizabeth	1 a. Person, enroll: enrollment b. Person, residence start: enrollment (London) c. Person, at place: enrollment (London)	8 a. Person, marry: wedding	14 a. Person, child born: birth	20 a. Person, observed: observation	28 a. Person, spouse dies: death	35 a. Person, move away from: move (London) b. Person, move to: move (New York) c. Person, residence stop: move (London) d. Person, residence start: move (New York)	42 a. Person, observed: observation
Elizabeth Residence at London	2 a. Residence start: enrollment			21 a. Residence, observed: observation		36 a. Residence stop: move	
Elizabeth Residence at New York						37 a. Residence start: move	43 a. Residence, observed: observation

Table 1: (continued)

States	Enrollment		Wedding		Birth		Observation @ 9		Death		Move		Observation @ 17
							Events						
Richard	3	a. Person, enroll: enrollment b. Person, residence start: enrollment (London) c. Person, at place: enrollment (London)	9	a. Person, Marry: wedding	15	a. Person, child born: birth	22	a. Person, observed: observation	29	a. Person stop: death b. Person, residence stop: death			
Richard Residence at London	4	a. Residence start: enrollment					23	a. Residence, observed: observation	30	a. Residence stop: death			
New York	5	a. Place, enroll: enrollment					24	a. Place, observed: observation			38	a. Place, person moves to: move b. Place residence start: move	44	a. Place, observed: observation
London	6	a. Place, enroll: enrollment b. Place, residence start: enrollment c. Place, person at: enrollment	10	a. Place, wedding: wedding	16	a. Place, child born: birth b. Place, residence start: birth	25	a. Place, observed: observation	31	a. Place, person dies: death b. Place, residence stop: death	39	a. Place, person moves from: move b. Place residence stop: move	45	a. Place, observed: observation

This example is implemented in a working database, see Sections A.1 and A.2.1.

In the New York – London example there are 45 intersections between states and events that contain meaningful influences, and these are listed and explained in detail in the cells of Table 1. For example, among the states influenced by Beatrice's birth are: 1) Beatrice's life as '*Person start: birth*', 2) Beatrice's residence at London as '*Residence start: birth'* and 3) both Elizabeth and Richard as '*Person, child born: birth*'. Section A.2.1 of the appendix describes in full detail all of the influences associated with Beatrice's birth.

3.5 The general temporal data model[8]

The *General Temporal Data Model* is simply the triad *Event* $\Leftrightarrow$ *Influence* $\Leftrightarrow$ *State* that allows one to associate states with events that influence them and with other states influenced by the same events.

The GTDM is capable of storing the history and time-evolving structure of an arbitrary collection of states at whatever level of detail required, more detail requires the definition of more states, events and influences. Additionally the structure of the GTDM naturally facilitates the generation of event lists describing the history of any type of state, and these lists of events are the basis of many types of longitudinal, survival and event history analysis.

There are many different ways to realize this abstraction and implement it in a working system, most of the variation having to do with how attributes of different types of states and events are conceptualized, stored and manipulated. Following is a detailed description of a suggested realization of the GTDM intended for implementation in a relational database. While reading the following sections (4 through 5) the reader may find it useful to consult the example databases that accompany this article. For details see the appendix Sections A.1 through A.4.

4. The structured population event history register

The *Structured Population Event History Register* is a relational database schema based on the GTDM. Although the GTDM is general and can record the related histories of any type of "thing", SPEHR is a GTDM realization adapted to record observations of the related histories of human beings, their residences at various locations and their memberships in various social groups. SPEHR retains the inherent benefits (schema-

---

[8] Data model – an abstract logical definition of the data that will be stored, including a detailed description of their structure and perhaps their behavior (Date, 2000).

invariant flexibility and conceptual integrity) of the GTDM but adds several features that are necessary in a working realization of the GTDM; namely the ability to manage and store different types of states and events in the same tables, and the ability to store and link the attributes of specific states and events. As a logical blueprint for a relational database SPEHR can be implemented using any standard relational database management system (Postgress, MySQL, MS SQL Server, IBM DB2, Oracle, MS Access etc.). The object of this work is to present the relational schema for SPEHR rather than provide details on implementing SPEHR in a specific relational database management system. Figure 3 contains an entity relationship diagram[9] of SPEHR that the reader may find useful while reading the following sections.

4.1 Metadata

The GTDM specifies three general objects – *states*, *events* and *influences* – that can each have many different *specific* types depending on the reality being represented by a GTDM-based database. SPEHR uses metadata, or data that describe other data, to allow the user to specify the domain of possible types for each of these general objects, and further to specify the unique specific type of each instance of the general objects.

SPEHR contains both metadata tables and general object tables. Metadata tables describe *types* of general objects, while each row in the general object table is a unique *instance* of the general object and must be associated with a specific type of that general object, as described in the metadata. This is accomplished by linking each row in the general object table to exactly one row in the associated metadata table, thereby identifying the specific type of each instance of the general object.

In the entity relationship diagram in Figure 3 the names of the metadata tables are suffixed with "_Types" to indicate that the table in question contains a list of the specific types that define the domain of types of a general object.

4.2 States

All three parts of the GTDM triad *Event ⇔ Influence ⇔ State* are stored in pairs of tables. One table defines the types of an object, while the other lists the specific instances of that object.

---

[9] Entity relationship diagram (ERD) – a diagram created during the design of a database to depict the modeled entities, their attributes, and their relationships to one another including directionality and cardinality. An ERD is a tool to help communicate the design and structure of a database (Anon).

Figure 2:   London – New York example: states tables

States	
State_ID	**State_Type_ID**
1	1
2	1
3	*2*
4	2
5	3
6	**2**
7	4
8	4
9	4
10	4
11	4
12	5

State_Types		
State_Type_ID	**State_Name**	**State_Description**
1	Place	
2	**Person**	
3	Union	
4	Residence	

State_Texts				
State_ID	**State_Type_ID**	**Attribute_Type_ID**	**Observation_Event_ID**	**State_Text**
1	1	1	1	London
2	1	1	1	New York
3	*2*	*1*	*1*	*Richard*
4	2	1	1	Elizabeth
5	3	1	2	Richard-Elizabeth Union
6	**2**	**1**	**2**	**Beatrice**
7	4	1	1	Richard Resident @ London
8	4	1	1	Elizabeth Resident @ London
9	4	1	2	Beatrice Resident @ London
10	4	1	3	Elizabeth Resident @ New York
11	4	1	3	Beatrice Resident @ New York

Let us use state to provide a concrete example of this concept, see Figure 2. As we saw before, all entities with a valid lifetime (e.g. people, marital unions, etc.) are considered states. There are different types of states, and our SPEHR example (Figure 1) contains four types of state: person, place, residence, and union. A list of these state types is stored in the **States_Types** table. A list of the specific states that correspond to the real people, places, residences and unions is stored in the **States** table which may contain many states of the same type, for example states of type person corresponding to Richard, Elizabeth and Beatrice.

In Figure 2 you can see that the state "Elizabeth" has a unique State_ID: 4. The state "Elizabeth" also has a State_Type_ID, 2, which when referenced with the **State_Types** table indicates that she is a state of type "person". (For a full discussion of Figure 2 and an explanation of the **State_Texts** table that it contains, see the appendix Section A.2.1)

SPEHR uses the **States** and **State_Types** tables to realize the states defined by the GTDM. The **States** table can contain many states of the same type (for example type 'person'), but each of these must correspond to a unique state of that type in the real world (for example, a unique person "Elizabeth"). To reflect that, each state must have a unique value in the State_ID field. The type of each state is specified by the value of the State_Type_ID (from the **State_Types** table) stored in each row.

4.3 Events

SPEHR uses the **Events** and **Event_Types** tables to realize the *events* defined by the GTDM. The **Events** table can contain many events of the same type (for example, type 'birth'), but each of these must correspond to a unique event of that type in the real world. To reflect that each will have a unique value in the Event_ID field. The type of each event is specified by the value of the Event_Type_ID (from the **Event_Types** table) stored in each row. *The Events table is the only table in SPEHR that includes a timestamp field; all dates and times are stored in this field of the Events table and nowhere else in the database.*

4.4 Influences

SPEHR uses the **Influences** and **Influence_Types** tables to realize the *influences* defined by the GTDM. The **Influences** table can contain many rows of the same type (for example, type 'Person start: birth'), but each of these must correspond to a unique influence of that type in the real world, and to reflect that each will have a unique

combination of values in the Event_ID (from the **Events** table), State_ID (from the **States** table) and Influence_Type_ID (from the **Influence_Types** table) fields.

The metadata contained in the **Influence_Types** table are the heart of SPEHR. They contain most of the critical information that defines the relationships a SPEHR database can represent, and as a result defining the metadata in the **Influence_Types** table is potentially difficult and must be thought through carefully.

Events can either start or stop states or they can occur sometime during a state. This temporal relationship to the state is part of the definition of an influence, through a link to the **Influence_Actions** table that stores the permissible Influence_Actions: 'start', 'stop', 'during', 'at beginning' and 'at end'. Some influences can occur only at the beginning or end of a state even though they do not themselves start or stop the state. An example is an influence that links a birth to the resulting newborn indicating that the newborn has started a residence membership state at some location. This influence must occur at the start of the new 'person' state but it is not the influence that properly starts the person state. Furthermore this influence is not required to signify the start of the new state (person) but is instead providing optional additional information signifying that the initial residence location of the newborn is known and recorded with a residence membership state that links the newborn to a specific location. States can have one and only one *start* or *stop* influence, and any other influences that properly link to the state at its beginning or end have the Influence_Action 'at beginning' or 'at end' rather than 'start' or 'stop'.

4.5 Attributes of Events and States

Because the three primary data tables presented so far – **States, Events** and **Influences** – can contain many different types of the general objects they store, they do not have attributes of their own to contain descriptive data whose type and definition can and does change depending on the type of object stored in each row. For example, a row storing a state of type *person* may have additional attributes to describe the person's name and sex; whereas a row storing a state of type *marital union* may not require any additional attributes, or a row storing a state of type *place* may require only one additional attribute in which to store the place's name. It is reasonable to assume that both states and events will need additional type-specific attributes, while influences will not because their type fully defines them.

Figure 3: Entity-relationship diagram of SPEHR
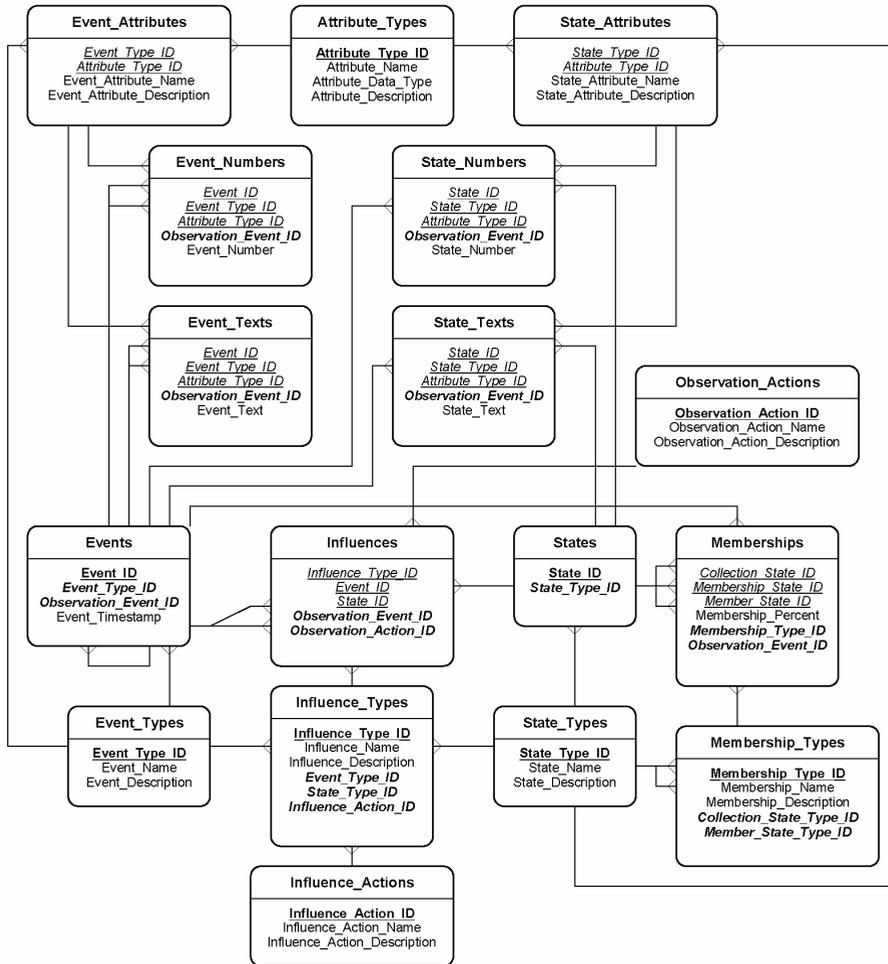
Tables are represented by boxes, table names are in the upper section of each box, field names are in the lower section of each box, primary keys are bold underlined, foreign keys are in bold italic, compound foreign keys that serve as primary keys are italic underlined, and relationships are represented by the "crows feet" lines between tables – with the many side of the relationship represented by the crow's foot.

The additional attributes of states and events are stored in a collection of separate tables and linked to the specific states and events that they describe. Three metadata tables contain descriptions of the types of attributes that can be stored by an individual SPEHR database. A general metadata table named **Attribute_Types** contains the domain of all possible attributes, including their names and data types. Many of these can be applied to both states and events so two more metadata tables named **State_Attributes** and **Event_Attributes** precisely define the domains of the attributes that can be applied to states and events. Each of these is linked to a fundamental attribute type in the metadata table **Attribute_Types**.

The attribute values themselves are stored in two collections of tables, one for states and one for events, with one table for each required data type[10]. The attribute values stored in these tables are linked directly to the states and events themselves. In the SPEHR schema displayed in Figure 3 there are four attribute tables – **State_Numbers**, **State_Texts, Event_Numbers** and **Event_Texts** – that store the individual attribute values that describe specific states and events. The State_ID or Event_ID of the specific state or event associated with a given attribute value is stored with the attribute value to link it to the state or event that it describes.

Many advantages and one disadvantage accrue from this approach to storing attribute values. The advantages include the ability to handle an arbitrary and dynamic set of attributes for states and events; and the fact that the metadata effectively comprise a data dictionary that is inextricably associated with the primary data, so much so that if it were removed the primary data would become meaningless. The first advantage is significant because it allows the definition and addition of arbitrary numbers of attributes either at design time or later *without making modifications to the database schema*. It allows legacy data to be retained and to retain their meaning without compromising the ability to store similar data with new definitions. Similarly the second advantage is significant in that the primary data (the attribute values themselves) cannot and will never be separated from the metadata that provide them with their meaning – *the data dictionary for a SPEHR database is built-in and will never be lost*.

The disadvantage is also important; namely that this way of storing attribute values makes their retrieval more difficult and less efficient. This is the only disadvantage to the SPEHR schema that the author views as potentially fatal, and so an experiment was conducted to test the efficiency of retrieving attributes from a SPEHR database storing a very large number of states (one million) and attribute values (four million). The experiment was conducted using the MS Access 2000 relational database management system implementing a simple version of the SPEHR schema with all relevant attribute tables properly indexed. The result conclusively demonstrated that arbitrary sets of

---

[10] In practice the database designer will decide which subset of the data types supported by the database management system is necessary for the SPEHR database being designed.

attributes can be retrieved for arbitrary single states and events almost instantaneously. Retrieving arbitrary sets of attributes for large numbers of states and events takes slightly longer – on order seconds – which is acceptable because "reports" of the type that require large numbers of states and events to be fully described do not need to be produced instantaneously.

4.6 Memberships

A prominent and often recurring theme in dynamic models of human populations is *membership*; a relationship that exists between a *collection* and a specific *member* of that collection. Collections (or aggregates) are simply groupings of individual members that persist from the time when there were at least two members until the group dissolves. Examples of collections include marital unions, families (including children), households (including non-family members), homesteads, villages and places. A place is a collection of the people who reside there; while the memberships themselves are more specifically *residences* at that location. Collections, members and memberships all have a beginning and an end, making them states in the SPEHR sense. Consequently, all three (collections, memberships and members) are stored in the **States** table in a SPEHR-based database. All that is necessary to record the special relationship that these states share with one another is to link them and label their relationship. This is accomplished by adding a table with three fields that stores the identifiers of the collection, membership and member states. An identifier of the membership type of the membership that links the three states is also needed.

SPEHR uses two tables named **Memberships** and **Membership_Types** to represent and store memberships. The **Memberships** table can contain many memberships of the same type (for example, type 'residence'), but each of these must correspond to a unique membership of that type in the real world, and to reflect that each will have a unique combination of values in the Collection_State_ID, Membership_State_ID and Member_State_ID fields. It is important to note that a collection and a member can have more than one membership relationship; each with a different membership state with its own unique identifier (and start and stop events/times).

4.7 Normalization

"Normalization" in database theory describes the formalized process of reducing or eliminating duplication of data in a database. A normalized database design is one in which very few or no facts are stored at more than one location in the database. Inconsistency or non-correspondence between multiple "copies" of the same fact is a serious problem, and it is one of the highest priorities of a database designer to prevent data corruption of this type.

A normalized design is one of the most important design criteria for SPEHR, in particular with respect to the temporal data stored in SPEHR. As a result the tables that comprise SPEHR do not duplicate any metadata or primary data. Critically, *the timestamps that provide the temporal dimension to SPEHR are stored in only one attribute of only one table* – the **Events** table – and it is through unique links to the **Events** table that data stored in other tables are given a temporal meaning.

4.8 Observations

A common goal of longitudinal studies is to relate the occurrence of an event to the accumulated exposure to that event. In population and health studies exposure is often measured in "person-years" and the object is to quantify the risk of some event controlling for the exposure to the event experienced by people of different types: differing ages or sexes for example. It is also critical to know for all units of analysis the first and last dates when they were observed, those being the dates before and after which nothing is known about them. For these and other reasons the "period of observation" for each unit of analysis is important for longitudinal analysis. Moreover from an operational point of view it is necessary to monitor and record when (and potentially where) study participants have been contacted or "observed".

Observation tracking has been built into the SPEHR schema. This is done through a special event type: "observation" in the **Event_Types** table. To support the linking of various pieces of primary data with the observation event at which they were recorded, an Observation_Event_ID field appears in many tables. The **Events** table itself has an Observation_Event_ID field to link each event to the observation event at which it was recorded. Likewise, the **Influences**, **Memberships**, **State_Numbers, State_Texts, Event_Numbers** and **Event_Texts** tables all have an Observation_Event_ID field to link each row in those tables to the observation event at which the data in the row were captured.

In addition to recording and linking observation events to primary data it is necessary to know for each state when periods of observation begin and end. For

example a state that describes the life of an individual person begins when the person is born and ends when they die, but the person may not be *observed* during their whole lifetime; observation may start with an enrollment event and end with an out-migration event or a death event. People and other entities can experience more complex patterns of repeated observation. It is clear that a range of different event types can initiate and/or terminate observation. Consequently the initiation and/or termination of observation should be an attribute of the influence that links a specific event to a specific state. To accommodate this there is an **Observation_Actions** table in SPEHR that functions as a metadata table containing a small number or rows that specify different observation actions: *start*, *stop-start*, *stop* and *no change*. Each influence stores an identifier corresponding to one of these observation actions to indicate what effect the influence has on the observation status of the state to which it is linked. "Start" indicates that observation is initiated, "stop" indicates that observation is terminated, "stop-start" indicates that observation is both stopped and restarted (this is used with observation events that both terminate the preceding period of observation and initiate the subsequent period of observation) and "no change" indicates that the influence has no effect on the observation status of the state. This representation of observation status provides a flexible yet simple way to explicitly specify the observation status of each state. This *greatly* facilitates the longitudinal analysis of data stored in a SPEHR database.

4.9 Event histories

Another element of longitudinal analysis that features in the design considerations of SPEHR is the fact that many longitudinal analysis techniques require "event histories" of some sort – essentially chronological lists of events that have occurred to each unit of analysis. Schema designs that distribute events and dates in many separate tables make it difficult to gather together all the events into a single view that displays an event history for each unit of analysis. It is in fact when attempting to create event histories that inconsistencies in duplicated dates often appear (see Section 4.7).

The idea of keeping all events in a single table is partly motivated by the necessity to generate event histories, and in practice it is indeed easy to generate event histories in a SPEHR database (see examples in Section A.4). Furthermore with the addition of two tables that define age intervals and calendar intervals (historical periods), it is then easy to generate age and historical period-specific "person-year" files that contain one row for each year that a person lived.

4.10 Metadata, data dictionaries and SPEHR database sharing

Together with its reflection of the GTDM, the metadata-driven concept that underlies SPEHR is what sets SPEHR apart from other temporal database designs. The GTDM provides it with a general and flexible way to represent temporal processes, and the metadata provide it with a general and flexible way to store that representation in a *schema-invariant* relational database. Schema-invariance refers to the fact that the logical structure of a SPEHR database does not need to change to expand and accommodate new entities. The important components of the metadata-driven approach are: 1) the metadata themselves and the meaning that they confer to the primary data (the data dictionary function), and 2) the fact that the metadata allow the database schema to remain constant while new entities are added and new meaning is given to the database.

The metadata-driven schema also allows different SPEHR database users to easily share their data with each other. This is possible because the underlying schemas of all SPEHR databases are the same and all that differentiates them are the specific metadata that each contains. Provided two SPEHR databases share some subset of metadata the primary data that those metadata describe can be pooled into one SPEHR database (that includes one extra table to store an identifier that differentiates the contributing data sources – see Section 5) and managed and/or analyzed as one dataset. All that is necessary to accomplish this is loading the primary data into a single SPEHR database that already contains their common (shared) subset of metadata.

To make this possible, someone must maintain a master archive of SPEHR metadata that serves as the standard metadata for all SPEHR databases that must be capable of interchanging and/or pooling data. The concept of a master metadata archive brings another advantage, namely it allows SPEHR "modules" developed by individual SPEHR users to be stored in the master archive from where they can be acquired by other users who wish to quickly and easily add a new module to their own SPEHR database. This facility could greatly expedite the extension and expansion of studies that wish to incorporate a complicated new module but do not have the time or resources to embark on a substantial database upgrade in order to manage the new data.

5. Multi-site structured population event history register

As research questions become more complex and seek measurements on less common events, it is natural to consider expanding single site studies to encompass multiple sites and hence multiple populations of larger sizes. Utilizing data from multiple sites, or better yet designing prospective multi-site studies, improves the representativeness of

the results and helps to combat the common criticism of small intensive longitudinal studies, namely that they produce "special" results that only pertain to the specific, small population that generated the data. These concerns suggest the need to combine data from several existing sites, or to initiate new research at a number of different sites, both of which require the interdigitation of potentially different data management systems or the creation of a new data management system that can handle data from all the sites simultaneously.

Another methodology currently being developed is "sample vital registration", conceived to substitute for full vital registration in parts of the world where full vital registration is lacking (MEASURE, 2006). Sample vital registration is essentially DSS-lite covering much larger geographic areas of a country and designed to cover enough of all parts of a country to produce representative data at a national level, but without much of the detail usually associated with a full DSS. Obviously this creates a requirement that data from several different regional populations be managed together. It is conceivable that as time progresses each region will want to add to and specialize the data that it collects in order to focus more finely on the issues that pertain to that region and not to others. Furthermore, a suggested enhancement to this methodology is to place one or more full DSSes in each region in order to provide the region with deeper, fuller longitudinal information on a smaller population that may be representative of the region, but not the larger area. The result of a combined DSS-sample vital registration system would be a comparatively inexpensive basic data collection platform that is both nationally representative and detailed enough at the regional level to provide a means of conducting in-depth investigations. Such a system would require a flexible data management facility that is able to incorporate heterogeneous data from different regions, different data collection systems and different historical periods and grow gracefully with the substantive requirements of the system for a long period of time.

As alluded to in Section 4.10, the potential for a SPEHR-based database to manage data from many sites easily and efficiently in one schema is one of SPEHR's significant advantages. The next section briefly introduces the multi-site version of SPEHR that is able to do that.

5.1 Multi-site SPEHR schema

In order to manage data from multiple sites or studies easily and efficiently, a straightforward addition to the existing single-site SPEHR schema is necessary. A table named **Sites** is added that contains a list of all the sites contributing data to the multi-site SPEHR database. Each site is assigned a unique identifier that is used to

differentiate each site's primary data from the primary data from other sites in all the primary data tables in SPEHR.

The metadata tables remain unchanged and contain exactly the same metadata as they do in the single-site version of SPEHR. It is when combining data from many sites that the metadata become critical, and the central metadata archive (see Section 4.10) plays a key role. Within each metadata table the unique identifiers associated with each row must be consistent across all SPEHR databases contributing data to the central multi-site SPEHR database. That is, the metadata that describe a specified type of a general object must have the same unique identifier in all of the SPEHR databases contributing data. That way when the data are all merged the metadata from each database will mean the same thing and provide the same meaning to the primary data coming from each database – obviously each metadata table row will occur only once in the combined database. For these reasons it is critical to manage the uniqueness and consistency of the metadata, and that is why the central metadata archive is important. Moreover it is unlikely that all the contributing databases will have the same metadata specification because they will have been individualized to some extent; the standardized metadata specification will allow the importation of only those primary data that are compatible and desired from the contributing databases.

6. Further components of SPEHR

Three major additional components would greatly enhance SPEHR's ability to support an efficient, reliable and useful production data management system for a longitudinal study of human populations. Ongoing work is addressing these:

1. a generalized metadata-driven facility to model and store both temporal and non-temporal integrity constraints on primary data,
2. a generalized metadata-driven facility to model and store questionnaire or other data collection instruments, and
3. a generalized suite of commonly required views and extraction tools that would be useful with reference to many or all different types of states stored in a SPEHR database.

The first is critical to maintaining the consistency and integrity of the data. Ordinary referential integrity constraints ensure that relationships between individual tables are maintained through the maintenance of primary and foreign key relationships, but they do little or nothing to provide a standardized easy way to ensure temporal integrity. Temporal integrity refers to a situation in which:

- all events are correctly sequenced,
- states that should not overlap in time do not,
- states that should overlap in time do,
- states that should abut or meet each other in time do, and
- states that should not "meet" do not, etc.

Essentially temporal integrity ensures that facts stored in a temporal database are: 1) associated with valid timestamps, 2) correctly sequenced chronologically, and 3) potentially associated with well-formed states that are in the correct sequence and relationship to other states and events in the database (see Date, Darwen and Lorentzos 2002a; 2002b). Because SPEHR is built around a clear conceptual standard and implemented in a static schema it will be possible to develop conceptually general, standard methods for assessing and enforcing temporal integrity.

The second component extends the usefulness of SPEHR in an important way and allows the joint, concurrent management of the data and the instruments that capture them together in one database. This allows the primary data to be permanently linked to the instrument and specific question that captured them, thus significantly extending the data dictionary to include the context in which data were collected. Managing the instruments and data together also allows the instruments to make use of the stored data to perform real-time validity checks on data that are being captured and to flag or reject potentially false or inconsistent data while it is being captured. Finally, such a system can be adapted to work on handheld computers that allow field workers to go into the field with true *smart* questionnaires that make full use of existing data to improve the quality of newly captured data and potentially save time in the capture process as well.

The third is not really a component on its own but rather an important sub-component of all components of SPEHR. SPEHR's generalized, static, metadata-driven schema makes it possible to define general, metadata-driven routines that actively interact with the database in many different ways. For example integrity checks of various types may make use of generalized routines that identify overlapping states of whatever type necessary, or miss-sequenced events, whether they belong to people, households or some other type of state. For analysis, general metadata-driven routines could be developed to calculate various measures of exposure. So instead of having to write separate routines to do similar things, it will be possible to write general routines that make use of SPEHR's metadata and static schema to accomplish the same thing with reference to many different types of the underlying objects. The person-year calculation described in the appendix Section A.4 is the first example of such a routine; it can calculate exposure for any type of state in exactly the same way: person-years, household-years, union-years etc.

These three additions are conceptualized within the same generalized metadata-driven framework to function in ways similar to the existing SPEHR schema and to interact with it at a deep level, sharing metadata and providing new metadata to enhance the integrity and meaning of data stored in the existing SPEHR schema.

7. Discussion

The work presented here is developed around the philosophy that generalization and standardization are greatly worth attaining in the service of improving efficiency, accuracy, reliability and comparability. To realize these aims it is necessary to develop basic, abstract representations of the real world and to identify the underlying similarities and congruencies among the entities we wish to manipulate, model, capture, store and retrieve. The GTDM provides the general abstract representation of temporal reality that we need to represent the interrelated histories of various entities. SPEHR provides a relational schema for implementing a working version of the GTDM in a relational database management system; one that is designed specifically to capture the interrelated histories of human beings. Building on that the multi-site version of SPEHR allows data describing the interrelated histories of people living in different populations and captured by different studies to be managed and manipulated together in one SPEHR-based database.

SPEHR is a flexible tool that allows a user to easily define the structured temporal data that they want to store and manipulate, to actually store and manipulate that data, and to refine and redefine the definitions of the data as time goes on – all without making changes to the schema of the relational database that implements SPEHR. As such SPEHR is not a "database for DSS" or a database for anything else in particular; to become a DSS database or a database for "X", a specific set of metadata must be defined and stored within SPEHR to allow it to store and manipulate the data collected by a DSS, or in the course of "X". It is worth noting again that the GTDM is sufficiently general to model the histories, interrelated or not, of many different kinds of "thing", not necessarily just people.

A common comment received from colleagues who have examined SPEHR is that the GTDM objects are so general that it is difficult to conceptualize how to reorganize the data into more familiar forms. Although this observation is cogent, it cuts two ways. Relational databases that store any meaningful level of temporal information describing human populations are all complex and take a lot of effort to understand and manipulate. Although the SPEHR schema is more abstract and perhaps more "tricky" than most, it has as its foundation a *few* simple, consistent concepts, and once those are mastered there is nothing else to learn about how to interact with a SPEHR database.

Once a user understands the basic idea they are set; they will not have to learn the concepts behind the next big revision of their database because the basic concepts behind a SPEHR database will not change.

Ongoing work addresses 1) the need to formalize integrity constraints in a general, metadata-driven way within SPEHR, 2) the need to develop and implement a general metadata-driven model of data capture instruments within the SPEHR schema, and 3) the need to develop general routines to support integrity checking, data manipulation, and extraction of highly manipulated data from SPEHR databases.

8. Acknowledgements

References

Allen, James F. 1983. "Maintaining Knowledge about Temporal Intervals." *Communications of the ACM*, 26(11 November 1983):832-43.

Allen, James F. and G. Ferguson. 1994. "Actions and Events in Interval Temporal Logic." *Journal of Logic and Computation*, 4(5):531-79.

Benzler, Justus and Samuel J. Clark. 2005. "Towards a Unified Timestamp with Explicit Precision." *Demographic Research*, 12(6):107-40.

Benzler, Justus, Kobus Herbst and Bruce MacLeod. 1998. "A Data Model for Demographic Surveillance Systems". http://www.indepth-network.org/publications /indepth_publications.htm. Accessed: 2006-01-12.

Clark, Samuel J. 2001. "Part 4: The Structured Population Event History Register - SPEHR." Pp. 356-78 in An Investigation into the Impact of HIV on Population Dynamics in Africa, Ph.D. dissertation in Demography. Philadelphia, Pennsylvania: University of Pennsylvania.

Date, C.J. 2000. "Chapter 1: An Overview of Database Management." Pp. 2-32 in An Introduction to Database Systems. Reading Massachusetts: Addison-Wesley.

Date, C.J., H. Darwen and N.A. Lorentzos. 2002a. "Chapter 11: Integrity Constraints I: Candidate Keys and Related Constraints." Pp. 187-212 in Temporal Data and the Relational Model. San Francisco: Morgan Kaufmann.

—. 2002b. "Chapter 12: Integrity Constrains II: General Constraints." Pp. 213-44 in Temporal Data and the Relational Model. San Francisco: Morgan Kaufmann.

—. 2002c. Temporal Data and the Relational Model. San Francisco: Morgan Kaufmann.

Etzion, Opher, Sushil Jajodia and Suryanarayana Sripada, Editors. 1998. Temporal Databases: Research and Practice, vol. 1399, Lecture Notes in Computer Science, Edited by G. Goos, J. Hartmanis and J. van Leeuwen. Berlin: Springer.

Gulutzan, P. and T. Pelzer. 1999. SQL-99 Complete, Really. Lawrence, Kansas: R&D Books.

HPTN. 2006. "HIV Prevention Trials Network - HVTN". http://www.hptn.org/. Accessed: 2006-02-01.

HVTN. 2006. "HIV Vaccines Trials Network - HVTN". http://www.hvtn.org/. Accessed: 2006-02-01.

IAVI. 2006. "International AIDS Vaccine Initiative - IAVI". http://www.iavi.org/. Accessed: 2006-02-01.

INDEPTH Network. 2004. "An International Network of Field Sites with Continuous Demographic Evaluation of Populations and Their Health in Developing Countries - INDEPTH". www.indepth-network.net; www.indepth-network.org. Accessed: 2006-10-05.

International Organization for Standardization. 2000. "ISO 8601:2000 Representation of Dates and Times." Geneva, Switzerland: International Organization for Standardization.

Jensen, Christian S. 2000. Temporal Database Management. Ph.D. dissertation in Department of Computer Science. Aalborg, Denmark: Aalborg University.

Jensen, Christian S., Curtis E. Dyreson, Michael Bohlen, James Clifford, Ramez Elmasri, Shashi K. Gadia, Fabio Grandi, Pat Hayes, Sushil Jajodia, Wolfgang Kafer*, et al.* 1998. "The Consensus Glossary of Temporal Database Concepts - February 1998 Version." in Temporal Databases: Research and Practice. Edited by O. Etzion, S. Jajodia and S. Sripada. Berlin: Springer.

MacLeod, B.B., J.F. Phillips and F.N. Binka. 1996. "Sustainable Software Technology Transfer: The Household Registration System." Pp. 302-10 in Encyclopedia of Library and Information Science, vol. 58. Edited by A. Kent. New York: Marcel Dekker.

MEASURE. 2006. "Sample Vital Registration with Verbal Autopsy (SAVVY)". http://www.cpc.unc.edu/measure/leadership/savvy.html. Accessed: 2006-02-02.

Phillips, James F, Bruce MacLeod and Brian Pence. 2000. "The Household Registration System: Computer Software for Rapid Dissemination of Demographic Surveillance Systems." *Demographic Research*, 2(6).

SAAVI. 2006. "South African AIDS Vaccine Initiative - SAAVI". http://www.saavi.org.za. Accessed: 2006-02-01.

Snodgrass, Richard T. 2000. Developing Time-Oriented Database Applications in SQL. San Francisco: Morgan Kaufmann Publishers.

Spaccapietra, S., C. Parent and E. Zimanyi. 1998. "Modeling Time from a Conceptual Perspective." Proceedings of *7th International Conference on Information and Knowledge Management*. Bethesda, Maryland. 2-7, November.

Appendix

A.1 Example databases

Accompanying this article via links on its startup page are example databases discussed in Sections A.2, A.3 and A.4. The databases are Access 2000 version MS Access databases, each in its own ".mdb" file. To access the databases you must have a version of MS Access that can open an Access 2000 version Access database. Each example database contains a number of tables and queries that can be opened and manipulated. Referential integrity relationships are also defined and can be viewed using the "relationship view" that is accessed by clicking on the relationship window icon ⊞ in the toolbar at the top of the Access main window.

A.2 Detailed examples

A.2.1 London – New York example

The example study described in Section 3.1 above and in Figure 1 is also implemented as a working example of the SPEHR schema in an accompanying MS Access 2000 database named "SPEHR-London-NewYork-2.0.mdb". The SPEHR schema presented in Figure 3 is implemented exactly and the metadata necessary to reflect the London – New York example study are inserted, including those presented in Table 1. The only difference is that real dates have been substituted for the simple time markers used in the example in Figure 1, and Richard and Elizabeth's birth events have been added to allow calculation of their ages. In addition to the tables and relationships necessary to implement the SPEHR schema there are a number of example queries (views written in SQL) defined in the example database to demonstrate that it is easy to view data in a SPEHR database in ways that are more familiar. The queries have self-explanatory names like: "Select_Person_Names_Sexes_Vital_Dates", "Select_Generation" etc. Select "Tables" or "Queries" from the "Objects" pane to the left of the database window within Access to view lists of the tables and queries in the database.

Figure 4 through Figure 7 display database tables from the London – New York example database. We will use these figures to examine how the information contained in the example in Figure 1 is actually stored in a SPEHR database, focusing on Beatrice's birth.

Examining the **States** table in Figure 4 reveals three records in which the value of the State_Type_ID field is '2' corresponding to the 'Person' state type as indicated in the **State_Types** table. These person states have State_IDs '3', '4', and '6'. Looking

up those State_IDs in the **State_Texts** table (Figure 5) retrieves three records that attach names to those state IDs: state '3' is Richard, state '4' is Elizabeth and state '6' is Beatrice. Furthermore we see that these state texts are all of attribute type '1' corresponding to 'name' and were observed at observation event '1' for Richard and Elizabeth, and observation event '2' for Beatrice. It is worth noting here that the IDs in a SPEHR database mean absolutely nothing beyond being unique identifiers, and the cardinal order of IDs is irrelevant.

Figure 5 displays the tables that link text attribute values to the states contained in the **States** table. The **Attribute_Types** table contains descriptions of the fundamental attribute types that can be utilized for either states or events and the **State_Attributes** table contains specific attribute types that apply to states, based on the fundamental attribute types stored in the **Attribute_Types** table. Finally the **State_Texts** table contains the specific text values that *are* the text-valued attributes attached to the states[11]. For example the actual text representing the names of Richard, Elizabeth and Beatrice are stored in the **State_Texts** table.

Turning to the **Events** table in Figure 6 we see two events related to Beatrice's birth: 1) the birth event itself with Event_ID '5' and Event_Type '3' which corresponds to the 'Birth' event type in the **Event_Types** table; and 2) the observation event that recorded the birth with Event_ID '2' and Event_Type '6' which corresponds to the 'Observation' event type in the **Event_Types** table.

---

[11] Refer to the entity relationship diagram in Figure 3 to see how the various IDs are inherited through these tables to maintain referential integrity.

Figure 4:   London – New York example: states tables

States	
State_ID	State_Type_ID
1	1
2	1
3	*2*
4	2
5	3
6	**2**
7	4
8	4
9	4
10	4
11	4
12	5

State_Texts				
State_ID	State_Type_ID	Attribute_Type_ID	Observation_Event_ID	State_Text
1	1	1	1	London
2	1	1	1	New York
3	*2*	*1*	*1*	*Richard*
4	2	1	1	Elizabeth
5	3	1	2	Richard-Elizabeth Union
6	**2**	**1**	**2**	**Beatrice**
7	4	1	1	Richard Resident @ London
8	4	1	1	Elizabeth Resident @ London
9	4	1	2	Beatrice Resident @ London
10	4	1	3	Elizabeth Resident @ New York
11	4	1	3	Beatrice Resident @ New York

State_Types		
State_Type_ID	State_Name	State_Description
1	Place	
2	**Person**	
3	Union	
4	Residence	

Figure 5:   London – New York example: state attributes tables

Attribute_Types			
Attribute_Type_ID	Attribute_Name	Attribute_Data_Type	Attribute_Description
1	**Name**	**Text**	
2	Sex	Number	

State_Attributes			
State_Type_ID	Attribute_Type_ID	State_Attribute_Name	State_Attribute_Description
1	1	Place name	
2	**1**	**Person name**	
2	2	Person sex	0 = Female, 1 = Male
3	1	Union name	
4	1	Residence name	
5	1	Process name	

Figure 5:   (continued)

				State_Texts
State_ID	State_Type_ID	Attribute_Type_ID	Observation_Event_ID	State_Text
1	1	1	1	London
2	1	1	1	New York
3	*2*	*1*	*1*	*Richard*
4	2	1	1	Elizabeth
5	3	1	2	Richard-Elizabeth Union
6	**2**	**1**	**2**	**Beatrice**
7	4	1	1	Richard Resident @ London
8	4	1	1	Elizabeth Resident @ London
9	4	1	2	Beatrice Resident @ London
10	4	1	3	Elizabeth Resident @ New York
11	4	1	3	Beatrice Resident @ New York
12	5	1	3	Process: Richard Death - Elizabeth & Beatrice Move

Altogether there are eight influence types related to Beatrice's birth in the **Influence_Types** table in Figure 7:

1. 'Place, child born: birth' with influence_Type_ID '6' that links a birth event to the location where the birth took place to simply indicate that a birth took place there regardless of whether a residence is also initiated at that location,

2. 'Person, child born: birth' with Influence_Type_ID '7' that links a birth event to the parents of the newborn,

3. 'Person start: birth' with Influence_Type_ID '8' that links a birth event to the newborn itself and properly *starts* the new person's life state,

4. 'Union, child born: birth' with Influence_Type_ID '9' that links the birth event to the union in which the birth occurred,

5. 'Person, residence start: birth' with Influence_Type_ID '23' that links a birth event to the newborn indicating that it has started a new residence state,

6. 'Residence start: birth' with Influence_Type_ID '24' that links a birth event to the residence state that has been started and properly starts the residence state,

7. 'Place, residence start: birth' with Influence_Type_ID '25' that links a birth event to the location where the residence begins indicating that a new residence has begun at that location, and

8. 'Person, observed: observation' with Influence_Type_ID '40' that links an observation event to a person.

Figure 6:   **London – New York example: events tables**

Event_Types		
Event_Type_ID	**Event_Name**	**Event_Description**
1	Enrollment	
2	Wedding	
3	**Birth**	
4	Death	
5	Move	
6	**Observation**	

Events			
Event_ID	**Event_Type_ID**	**Observation_Event_ID**	**Event_Timestamp**
1	1	1	1990-03-20 12:00:00 PM
2	**6**	**2**	**1993-11-09 12:00:00 PM**
3	6	3	1997-08-18 12:00:00 PM
4	2	2	1991-04-15 12:00:00 PM
5	**3**	**2**	**1992-10-14 12:00:00 PM**
6	4	3	1995-05-11 12:00:00 PM
7	5	3	1996-08-23 12:00:00 PM
9	3	1	1965-08-21 12:00:00 PM
10	3	1	1967-02-15 12:00:00 PM

Figure 7:   London – New York example: influences tables

Influence_Actions		
Influence_Action_ID	**Influence_Action_Name**	**Influence_Action_Description**
1	**Start**	**Influence starts a State**
2	**During**	**Influence is during a State**
3	Stop	Influence stops a State
4	**At beginning**	**Influence at the beginning of a State**
5	At end	Influence at the end of a State

Influence_Types					
Influence_Type_ID	**Influence_Name**	**Influence_Description**	**Event_Type_ID**	**State_Type_ID**	**Influence_Action_ID**
1	Place, enroll: enrollment		1	1	2
2	Person, enroll: enrollment		1	2	2
3	Place, wedding: wedding		2	1	2
4	Person, marry: wedding		2	2	2
5	Union start: wedding		2	3	1
6	Place, child born: birth		3	1	2
7	*Person, child born: birth*		*3*	*2*	*2*
8	**Person start: birth**		**3**	**2**	**1**
9	Union, child born: birth		3	3	2
11	Place, person dies: death		4	1	2
12	Person, spouse dies: death		4	2	2
13	Person, parent dies: death		4	2	2
14	Person stop: death		4	2	3
15	Union stop, spouse dies: death		4	3	3
16	Place, person moves from: move		5	1	2
17	Place, person moves to: move		5	1	2
18	Person, move away from: move		5	2	2
19	Person, move to: move		5	2	2
20	Person, residence start: enrollment		1	2	2
21	Residence start: enrollment		1	4	1
22	Place, residence start: enrollment		1	1	2
23	**Person, residence start: birth**		**3**	**2**	**4**

Figure 7:   (continued)

Influence_Types					
Influence_Type_ID	**Influence_Name**	**Influence_Description**	**Event_Type_ID**	**State_Type_ID**	**Influence_Action_ID**
24	Residence start: birth		3	4	1
25	Place, residence start: birth		3	1	2
26	Person, residence stop: death		4	2	5
27	Residence stop: death		4	4	3
28	Place, residence stop: death		4	1	2
29	Person, residence stop: move		5	2	2
30	Residence stop: move		5	4	3
33	Place, residence stop: move		5	1	2
34	Person, residence start: move		5	2	2
35	Residence start: move		5	4	1
36	Place, residence start: move		5	1	2
39	Place, observed: observation		6	1	2
40	**Person, observed: observation**		**6**	**2**	**2**
41	Union, observed: observation		6	3	2
42	Residence, observed: observation		6	4	2
47	Place, person at: enrollment		1	1	2
48	Person, at place: enrollment		1	2	2

Observation_Actions		
Observation_Action_ID	**Observation_Action_Name**	**Observation_Action_Description**
1	**Start**	
2	**Stop-Start**	
3	Stop	
4	**No change**	

Figure 7: (continued)

		Influences		
Influence_Type_ID	Event_ID	State_ID	Observation_Event_ID	Observation_Action_ID
1	1	1	1	1
1	1	2	1	1
2	1	3	1	1
2	1	4	1	1
3	4	1	2	4
4	4	3	2	4
4	4	4	2	4
5	4	5	2	1
6	5	1	2	4
7	5	3	2	4
7	5	4	2	4
8	5	6	2	1
8	9	3	1	4
8	10	4	1	4
9	5	5	2	4
11	6	1	3	4
12	6	4	3	4
13	6	6	3	4
14	6	3	3	3
15	6	5	3	3
16	7	1	3	4
17	7	2	3	4
18	7	4	3	4
18	7	6	3	4
19	7	4	3	4
19	7	6	3	4
20	1	3	1	4
20	1	4	1	4
21	1	7	1	1
21	1	8	1	1
22	1	1	1	4
23	5	6	2	4
24	5	9	2	1

Figure 7: (continued)

		Influences		
Influence_Type_ID	Event_ID	State_ID	Observation_Event_ID	Observation_Action_ID
25	5	1	2	4
26	6	3	3	4
27	6	7	3	3
28	6	1	3	4
29	7	4	3	4
29	7	6	3	4
30	7	8	3	3
30	7	9	3	3
33	7	1	3	4
34	7	4	3	4
34	7	6	3	4
35	7	10	3	1
35	7	11	3	1
36	7	2	3	4
39	2	1	2	2
39	2	2	2	2
39	3	1	3	2
39	3	2	3	2
40	2	3	2	2
40	2	4	2	2
40	**2**	**6**	**2**	**2**
40	3	4	3	2
40	3	6	3	2
41	2	5	2	2
42	2	7	2	2
42	2	8	2	2
42	2	9	2	2
42	3	10	3	2
42	3	11	3	2
45	6	12	3	1
46	7	12	3	3
47	1	1	1	4
48	1	3	1	4
48	1	4	1	4

The event type IDs associated with these influence types are '3' corresponding to 'Birth' for the first seven and '6' corresponding to 'Observation' for the last. The state types associated with these influence types are '1', '2', '2', '3', '2', '4', '1' and '2' corresponding to 'Place', 'Person', 'Person', 'Union', 'Person', 'Residence', 'Place' and 'Person'. The reader can examine the influence action types associated with each influence type and confirm that they are appropriate.

Turning to the influences themselves there are nine influences associated with Beatrice's birth. Examining them in the order they appear in the Influences table in Figure 7:

1. the first links the birth event (Event_ID '5') with state '1' (London) and is of influence type '6' ('Place, child born: birth') indicating that a birth occurred at London,

2. the second links the birth event (Event_ID '5') with state '3' (Richard) and is of influence type '7' ('Person, child born: birth') indicating that a child was born to Richard,

3. the third is the same as the second except it links the birth event to Elizabeth,

4. the fourth with influence type '8' ('Person start: birth') links the birth to state '6' (Beatrice) as the start of her life,

5. the fifth with influence type '9' ('Union, child born: birth') links the birth to state '5' (the union between Richard and Elizabeth) to indicate that their union gave rise to Beatrice's birth,

6. the sixth with influence type '23' ('Person, residence start: birth') links the birth to Beatrice in another way, this time indicating that Beatrice has started a residence episode,

7. the seventh with influence type '24' ('Residence start: birth') links the birth to state '9' (Beatrice's residence at London) to start Beatrice's residence state at London,

8. the eighth with influence type '25' ('Place, residence start: birth') links the birth to London in another way, this time indicating that a new residence state has been started at London, and

9. the ninth and last influence with influence type '40' ('Person, observed: observation') links the observation event '2' to Beatrice indicating that this was the observation event at which her birth was recorded (this observation event ID is also recorded in the Observation_Event_ID field of all the influences associated with Beatrice's birth).

The last field in the **Influences** table in Figure 7 is of special interest. This field contains the ID of the observation action associated with each influence. The various influences associated with Beatrice's birth are associated with observation actions '1', '2' and '4' corresponding to "start", "stop-start" and "no change". The two influences that link the birth event to Beatrice's parents are associated with the "no change" observation action to indicate that Beatrice's birth does not affect the observation status of her parents. The influence linking the birth to Beatrice as the start of her life is linked to the "start" observation action to signify that her birth initiates observation for Beatrice (because her birth occurred during the study period). The influence linking the birth to Beatrice indicating that she is starting a residence state does not affect Beatrice's observation status and thus is linked to observation action "no change". Finally the influence linking the observation event to Beatrice is associated with observation action "stop-start" which signifies that the observation event terminates the earlier period of observation and initiates the next period of observation – this is important because only *complete* (i.e. closed or terminated) periods of observation are included in analyses. At this point the reader may be wondering about the observation status of the residence state initiated by Beatrice's birth; there is another influence associated with the "start" observation action that correctly sets the observation status for the residence state, and it is left to the reader to locate the relevant residence state and its associated influences. Finally, the reader may want to examine Richard's death and trace through all of the influences linked to that event and how they relate various states to Richard's death.

A.2.2 SPEHR examples using familiar longitudinal data collection schemes

There are many longitudinal data collection mechanisms that produce temporal data in a variety of formats. It is the central thesis of this work that longitudinal data, no matter how they are collected or generated, have the same underlying structure and can thus be conceptualized, stored and managed in a single standard way, namely by invoking the GTDM. This section will discuss how two common longitudinal data collection schemes can be SPEHRized: the "repeated round" demographic survey and the population register.

A.2.2.1 "Repeated Round" demographic survey

The "repeated round" demographic survey is the data collection method preferred by most DSS sites. Each household or individual is visited repeatedly on a regular

schedule or "round" once every 3 or 4 months, or in some cases once per year. At each of these regular visits information is collected about the status of each study participant, and in particular information relating to the arrival or departure of study participants is recorded – data that describe births and in-migrations that add new study participants and deaths and out-migrations that remove existing study participants. The data collected in this fashion fully describe the study population at the conclusion of each round[12]. These data are typically recorded in a household register or similar book organized around villages, compounds and households and consisting of one "line" per household member per round. The identifier associated with individuals is typically a concatenation of their village, household and line numbers that forms a three-part ID linking them to their village and household (and sometimes a compound as well). As mentioned above in relation to the HRS, this poses problems when people move between villages and households and thereby change their identifiers.

---

[12] Other round-specific data are also collected at each (but not every) round, while the core demographic data that describe the study population are collected at every round and account for all members of the study population at each round.

Figure 8: **Example data for "Repeated Round" demographic survey**

Village #	HH #	Line #	Name	Interview Date	Observation	Sex	DOB	DOD	Date of In-migration	Date of Out-migration	Date of Parturition	Mother ID (V#.HH#.Line#)	Father ID (V#.HH#.Line#)
1	1	1	V1.HH1.P1	2003-01-05	Enroll	Male	1950-02-12						
1	1	1	V1.HH1.P1	2003-05-05	Continue								
1	1	1	V1.HH1.P1	2003-09-02	Continue								
1	1	1	V1.HH1.P1	2003-12-31	Continue								
1	1	2	V1.HH1.P2	2003-01-05	Enroll	Female	1952-11-02						
1	1	2	V1.HH1.P2	2003-05-05	Continue								
1	1	2	V1.HH1.P2	2003-09-02	Continue								
1	1	2	V1.HH1.P2	2003-12-31	Died			2003-11-02					
1	1	3	V1.HH1.P3	2003-01-05	Enroll	Male	1976-05-18						
1	1	3	V1.HH1.P3	2003-05-05	Continue								
1	1	3	V1.HH1.P3	2003-09-02	Continue								
1	1	3	V1.HH1.P3	2003-12-31	Continue								
1	1	4	V1.HH1.P4	2003-01-05	Enroll	Female	1980-07-25						
1	1	4	V1.HH1.P4	2003-05-05	Continue								
1	1	4	V1.HH1.P4	2003-09-02	Continue								
1	1	4	V1.HH1.P4	2003-12-31	Continue								
1	2	1	V1.HH2.P1	2003-05-05	In/out mig	Male	1977-01-02		2003-02-14	2003-03-01			
1	2	1	V1.HH2.P1	2003-09-02	In-mig				2003-07-03				
1	2	1	V1.HH2.P1	2003-12-31	Continue/Birth						2003-10-19		
1	2	2	V1.HH2.P2	2003-05-05	In-mig	Female	1981-05-12		2003-02-14				
1	2	2	V1.HH2.P2	2003-09-02	Continue								
1	2	2	V1.HH2.P2	2003-12-31	Gave Birth						2003-10-19		
1	2	3	V1.HH2.P3	2003-12-31	Born	Female	2003-10-19					1.2.2	1.2.1

Figure 8:   **(continued)**

Village #	HH #	Line #	Name	Interview Date	Observation	Sex	DOB	DOD	Date of In-migration	Date of Out-migration	Date of Parturition	Mother ID (V#.HH#.Line#)	Father ID (V#.HH#.Line#)
2	1	1	V2.HH1.P1	2003-05-05	In-mig	Male	1985-08-09		2003-03-11				
2	1	1	V2.HH1.P1	2003-09-02	Continue								
2	1	1	V2.HH1.P1	2003-12-31	Out-mig/Birth					2003-12-20	2003-12-06		
2	1	2	V2.HH1.P2	2003-05-05	In-mig	Female	1986-02-13		2003-03-11				
2	1	2	V2.HH1.P2	2003-09-02	Continue								
2	1	2	V2.HH1.P2	2003-12-31	Out-mig/Gave Birth					2003-12-20	2003-12-06		
2	1	3	V2.HH1.P3	2003-12-31	Out-mig/Born	Male	2003-12-06			2003-12-20		2.1.2	2.1.1
2	2	1	V2.HH2.P1	2003-01-05	Enroll	Male	1938-04-03						
2	2	1	V2.HH2.P1	2003-05-05	Continue								
2	2	1	V2.HH2.P1	2003-09-02	Continue								
2	2	1	V2.HH2.P1	2003-12-31	Died			2003-09-12					
2	2	2	V2.HH2.P2	2003-01-05	Enroll	Female	1940-11-16						
2	2	2	V2.HH2.P2	2003-05-05	Continue								
2	2	2	V2.HH2.P2	2003-09-02	Died			2003-06-01					
2	2	3	V2.HH2.P3	2003-01-05	Enroll	Female	1960-07-17						
2	2	3	V2.HH2.P3	2003-05-05	Continue								
2	2	3	V2.HH2.P3	2003-09-02	Continue								
2	2	3	V2.HH2.P3	2003-12-31	Continue								
2	2	4	V2.HH2.P4	2003-01-05	Enroll	Female	1962-11-03						
2	2	4	V2.HH2.P4	2003-05-05	Continue								
2	2	4	V2.HH2.P4	2003-09-02	Continue								
2	2	4	V2.HH2.P4	2003-12-31	Continue								

Figure 8 contains example data organized in the manner typical of a "repeated round" demographic survey. In this example there are two villages ('1' and '2'), two households ('1' and '2') per village and various numbers of people within each household (and hence various line numbers). There are four observation rounds with observations on January 5, 2003; May 5, 2003; September 2, 2003 and December 31, 2003. Each participant has one row for each round listing the observation date and the observation outcome. The possible observation outcomes include born, died, in-migration, out-migration and gave birth, with columns to contain the dates of these events and the IDs of the mother and father when a newborn is added to the register. Additionally when a participant is first captured by the system their sex and date of birth are recorded.

Take for example the individual imaginatively named 'V1.HH2.P1'. This person was first captured in round two on May 5, 2003 when he was observed to have both in-migrated and out-migrated during the previous inter-round period. He was recorded as 'male', having been born on January 2, 1977, having in-migrated to the study area on February 14, 2003 and having out-migrated again on March 1, 2003. This information was provided by his wife 'V1.HH2.P2' who remained in the study area and was interviewed on May 5, 2003. V1.HH2.P1 subsequently in-migrates again on July 3, 2003 and is interviewed on September 2, 2003. On October 19, 2003 V1.HH2.P1's wife V1.HH2.P2 gave birth to their child V1.HH2.P3, and all of them were interviewed again on December 31, 2003.

The example presented in Figure 8 is also implemented as a working example of SPEHR using MS Access 2000 in a database file named "SPEHR-Repeated-Demographic-Survey-OR-Population-Register-2.0.mdb" that accompanies this article. For simplicity and clarity the SPEHR implementation designed to store the example "repeated round" data focuses exclusively on people without taking into account villages or households. To add these to the model metadata need to be created to define villages, households, memberships of individuals in households, residences of individuals at villages and perhaps residences of households at villages. Thinking through these additions is left to the reader (see the residence membership in the preceding London – New York example database to get started). Like the London – New York example database the "repeated round" example database contains the standard SPEHR tables and the specific metadata necessary to store the "repeated round" data, the same example analytic queries as the London – New York example and several additional example analytic queries specific to the "repeated round" example.

Figure 9: "Repeated Round" example: states tables

States	
State_ID	State_Type_ID
3	2
4	2
5	2
6	2
8	2
10	2
13	2
14	2
15	2
17	2
18	2
19	2
20	2

State_Texts				
State_ID	State_Type_ID	Attribute_Type_ID	Observation_Event_ID	State_Text
3	2	1	1	V1.HH1.P1
4	2	1	1	V1.HH1.P2
5	2	1	1	V1.HH1.P3
6	2	1	1	V1.HH1.P4
8	**2**	**1**	**2**	**V1.HH2.P1**
9	2	1	2	V1.HH2.P2
10	2	1	4	V1.HH2.P3
13	2	1	2	V2.HH1.P1
14	2	1	2	V2.HH1.P2
15	2	1	3	V2.HH1.P3
17	2	1	1	V2.HH2.P1
18	2	1	1	V2.HH2.P2
19	2	1	1	V2.HH2.P3
20	2	1	1	V2.HH2.P4

State_Types		
State_Type_ID	State_Name	State_Description
1	Place	A defined location on the surface of Earth
2	Person	A human being
3	Household	A collection of related human beings

Refer to the example database "SPEHR-Repeated-Demographic-Survey-OR-Population-Register-2.0.mdb" for a list of the State_Numbers that contain attribute values for the village, household and line numbers associated with each person in the "repeated rounds" example – the table is too large to reproduce here.

To demonstrate how the "repeated rounds" data are stored in a SPEHR database we will examine the metadata necessary to configure the "repeated rounds" SPEHR with specific attention to the records that describe our example individual V1.HH2.P1.

The **States** table in Figure 9 contains a list of the 'Person' states corresponding to the individuals in the "repeated rounds" example with state '8' corresponding to individual V1.HH2.P1. Figure 10 displays the events present in the "repeated rounds" example and of particular relevance to individual V1.HH2.P1 are events of type 'Observation', 'Birth' and 'Move'. In the Events table we see eight events relevant to V1.HH2.P1: three observation events, two birth events, and three move events. There are eight influences relevant to V1.HH2.P1 in Figure 11, these are: three observation

influences of influence type '5' associated with the three observations that affect V1.HH2.P1 in rounds 2-4; two birth events, one that starts his life and is associated with an influence of type '7', and one that links him to the birth of his daughter V1.HH2.P3 and is associated with an influence of type '8'; two 'Move, brings into study area, person' influences of type '10' that link V1.HH2.P1 to the move events that in-migrated him twice into the study area; and one 'Move, removes person from study area, person' influence of type '11' that links V1.HH2.P1 to the move event that removed him from the study area.

Studying the observation actions associated with the influences relevant to V1.HH2.P1 reveals that the observation influences are all associated with "stop-start" observation actions (for the same reason as in the London – New York example); both birth influences are associated with observation action "no change" because his birth is simply there to fix his birth date so that his age can be calculated, and his daughter's birth does not affect his observation status; the two in-migration influences are associated with "start" observation actions signifying that V1.HH2.P1 came under observation both times that he moved into the study area; and finally the out-migration influence is associated with the "stop" observation action indicating that V1.HH2.P1 was no longer observed after he moved out of the study area.

A.2.2.2 Population register

A population register is a special form of a "repeated round" demographic survey with the key difference that the population is monitored continuously and events are recorded as they happen instead of waiting for a special scheduled observation event to capture them. In this case there are "registration" events that occur as soon as possible after an event of interest and conceptually correspond to the "observation" events that have been described in the previous examples, with the key difference that they are not scheduled and do not happen at regular intervals but instead only when they are necessary to capture an event of interest.

The preceding examples suggest that the underlying structure of longitudinal information is the same no matter how it is collected or how the data are initially recorded directly following collection. A population register is implemented in SPEHR in exactly the same way as the preceding two examples; the relevant states, events and influences are defined by metadata and specific instances of these are recorded in the database. In fact the SPEHR implementation of a population register is logically identical to the SPEHR implementation of the "repeated round" demographic survey with the only difference being that observation events take place at arbitrary times as soon after the relevant events as possible.

Figure 10: "Repeated Rounds" example: events tables

Event_Types		
Event_Type_ID	**Event_Name**	**Event_Description**
1	Initiate	Initiate a study
2	**Observation**	**Observe something**
3	**Birth**	**A person is born**
4	Death	A person dies
5	**Move**	**A person moves**

Events			
Event_ID	**Event_Type_ID**	**Observation_Event_ID**	**Event_Timestamp**
1	1	1	2003-01-05 12:00:00 PM
2	**2**	**2**	**2003-05-05 12:00:00 PM**
3	**2**	**3**	**2003-09-02 12:00:00 PM**
4	**2**	**4**	**2003-12-31 12:00:00 PM**
5	3	1	1950-02-12 12:00:00 PM
6	3	1	1952-11-02 12:00:00 PM
7	3	1	1976-05-18 12:00:00 PM
8	3	1	1980-07-25 12:00:00 PM
9	**3**	**1**	**1977-01-02 12:00:00 PM**
10	3	1	1981-05-12 12:00:00 PM
11	**3**	**4**	**2003-10-19 12:00:00 PM**
12	3	1	1985-08-09 12:00:00 PM
13	3	1	1986-02-13 12:00:00 PM
14	3	4	2003-12-06 12:00:00 PM
15	3	1	1938-04-03 12:00:00 PM
16	3	1	1940-11-16 12:00:00 PM
17	3	1	1960-07-17 12:00:00 PM
18	3	1	1962-11-03 12:00:00 PM
19	4	4	2003-11-02 12:00:00 PM
20	4	4	2003-09-12 12:00:00 PM
21	4	3	2003-06-01 12:00:00 PM
22	**5**	**2**	**2003-02-14 12:00:00 PM**
23	5	2	2003-03-11 12:00:00 PM
24	5	4	2003-12-20 12:00:00 PM
25	**5**	**2**	**2003-03-01 12:00:00 PM**
26	**5**	**3**	**2003-07-03 12:00:00 PM**

Figure 11: "Repeated Rounds" example: influences tables

Influence_Actions		
Influence_Action_ID	Influence_Action_Name	Influence_Action_Description
1	Start	Influence starts a State
2	During	Influence is during a State
3	Stop	Influence stops a State
4	At beginning	Influence at the beginning of a State
5	At end	Influence at the end of a State

Observation_Actions		
Observation_Action_ID	Observation_Action_Name	Observation_Action_Description
1	Start	
2	Stop-Start	
3	Stop	
4	No change	

Influence_Types					
Influence_Type_ID	Influence_Name	Influence_Description	Event_Type_ID	State_Type_ID	Influence_Action_ID
1	Initiate, enrolls, place	Initiation of study triggers enrollment of a study area	1	1	2
2	Initiate, enrolls, person	Initiation of study triggers enrollment of a person	1	2	2
3	Initiate, enrolls, household	Initiation of study triggers enrollment of a household	1	3	2
4	Observation, observes, place	Observation of a place	2	1	2
5	**Observation, observes, person**	**Observation of a person**	**2**	**2**	**2**
6	Observation, observes, household	Observation of a household	2	3	2
7	**Birth, starts, person**	**Birth begins a person's life**	**3**	**2**	**1**

Figure 11:   (continued)

Influence_Types					
Influence_Type_ID	Influence_Name	Influence_Description	Event_Type_ID	State_Type_ID	Influence_Action_ID
8	**Birth, happens to, person**	**Birth happens to a person**	**3**	**2**	**2**
9	Death, stops, person	Death ends a person's life	4	2	3
10	**Move, brings into study area, person**	**Move brings someone into the study area**	**5**	**2**	**2**
11	**Move, removes from study area, person**	**Move takes someone out of the study area**	**5**	**2**	**2**
12	Move, brings into study area, household	Move brings a household into the study area	5	3	2
13	Move, removes from study area, household	Move takes a household out of the study area	5	3	2

Influences				
Influence_Type_ID	Event_ID	State_ID	Observation_Event_ID	Observation_Action_ID
2	1	3	1	1
2	1	4	1	1
2	1	5	1	1
2	1	6	1	1
2	1	17	1	1
2	1	18	1	1
2	1	19	1	1
2	1	20	1	1
5	2	3	2	2
5	2	4	2	2
5	2	5	2	2
5	2	6	2	2
5	**2**	**8**	**2**	**2**

Figure 11:   (continued)

			Influences	
Influence_Type_ID	Event_ID	State_ID	Observation_Event_ID	Observation_Action_ID
5	2	9	2	2
5	2	13	2	2
5	2	14	2	2
5	2	17	2	2
5	2	18	2	2
5	2	19	2	2
5	2	20	2	2
5	3	3	3	2
5	3	4	3	2
5	3	5	3	2
5	3	6	3	2
5	**3**	**8**	**3**	**2**
5	3	9	3	2
5	3	13	3	2
5	3	14	3	2
5	3	17	3	2
5	3	19	3	2
5	3	20	3	2
5	4	3	4	2
5	4	5	4	2
5	4	6	4	2
5	**4**	**8**	**4**	**2**
5	4	9	4	2
5	4	10	4	2
5	4	19	4	2
5	4	20	4	2
7	5	3	1	4
7	6	4	1	4
7	7	5	1	4
7	8	6	1	4
7	**9**	**8**	**1**	**4**
7	10	9	1	4

Figure 11:   (continued)

Influence_Type_ID	Event_ID	State_ID	Observation_Event_ID	Observation_Action_ID
			Influences	
7	11	10	4	1
7	12	13	1	4
7	13	14	1	4
7	14	15	4	1
7	15	17	1	4
7	16	18	1	4
7	17	19	1	4
7	18	20	1	4
8	11	9	4	4
8	14	14	4	4
8	**11**	**8**	**4**	**4**
8	14	13	4	4
9	19	4	4	3
9	20	17	4	3
9	21	18	3	3
10	**22**	**8**	**2**	**1**
10	22	9	2	1
10	23	13	2	1
10	23	14	2	1
10	**26**	**8**	**3**	**1**
11	24	13	4	3
11	24	14	4	3
11	24	15	4	3
11	**25**	**8**	**2**	**3**

If one needs to add or link to other descriptive data such as the individual-level information contained in a social security index, there are two options. The first is to store that data directly as attributes of the individuals, and the second is to simply store each individual's social security number as an attribute that can be used to link to the social security index when necessary. The second is the preferred method since the external data source is likely to be updated from time to time.

A.3 Multi-site SPEHR examples: London – New York and Johannesburg – Durban

As with the single-site version of SPEHR a working example of the multi-site version has been created and accompanies this article as an MS Access 2000 database named "SPEHR-Merged-2.0.mdb". The sites that contribute to this example database are the "London – New York" example site discussed in Sections 3.1 and A.2.1 and displayed in Figure 1 and another example site called "Joburg – Durban". The Joburg – Durban example is very similar to the London – New York example, differing only in names and dates. An example single-site SPEHR database of the Johannesburg – Durban example also accompanies this work named "SPEHR-Joburg-Durban-2.0.mdb". The contents of the London – New York and Joburg – Durban single-site SPEHR databases are combined according to the procedure described above in 5.1 to yield the multi-site example database.

A.4 Analysis & data extraction from a SPEHR database

To be stored in SPEHR longitudinal data are split apart into abstract components that are recorded in different tables in the database. The resulting states, events and influences bear little resemblance to the everyday things that are described by the data. Because all SPEHR databases are relational the structured query language (SQL) is used to manipulate and extract data from the database. The example databases that accompany this article include a number of SQL queries (or views) that join the various elements of the data back together into familiar-looking entities, like people. There are also queries that when executed in sequence calculate person-years exposed, fertility rates and mortality rates in arbitrary user-defined time intervals and age groups. The time intervals and age groups are defined in tables that can be edited by the user, **AD_Time_Intervals** and **AD_Age_Groups**. To closely examine the queries select the "Queries" tab in the "Objects" pane to the left of the database main window. Double click on any of the listed queries to run them and examine the results; right click and select "design view" to examine their design and the SQL code that they contain (a further right click and selecting "SQL view" reveals the SQL). Finally, click on the "Modules" tab in the "Objects" pane to reveal the "Calculations" module. Within the Calculations module you will find a function named "ftnMakeRates". This is a straightforward function that simple executes a number of queries in sequence to form observed intervals, correct for any overlapping observed intervals, and calculate period-age-sex-specific person-years and rates. The code is commented to indicate the purpose of each query and should be self explanatory. To run this function and calculate

person-years and demographic rates select the "Forms" tab in the "Objects" pane and double click on the "Calculations" form to open it. This will immediately do the calculations for you using the current values in the time intervals and age groups tables and then display a button labeled "Calculate Rates" that you can push to repeat the calculation with either the same or new values in the time intervals and age groups tables. The results are always written to the **A_Time_Sex_Age_Rates** table.

A.4.1 London – New York example analytic queries

Figure 12 displays various views from the London – New York example. The first displays a list of the people stored in the database together with their ID, name, sex and vital dates. This should look familiar and comfort those who are at this point nervous about the abstract way in which SPEHR stores data (this query can be run in real time by double-clicking on its name in the example database). The second example query lists the "generations" present in the London – New York example; this query essentially links parents with children and displays their relationship in a single record. Beatrice is the only person born in the example, and you can see that she is linked properly to both her parents. This type of reconstruction is obviously critical to any study of fertility. Last there is a query that lists the "event histories" of all the people in the example. This is a more abstract view of the data but very useful; it displays a chronological list of every recorded happening of every person. This list can be further restricted to include only certain types of events or events that have certain impacts – like a specific change in observation status – and can then be manipulated further to yield interesting new views of the data, for example interbirth intervals.

Figure12:   London – New York example: various analysis queries

Query: Select_Person_Names_Sexes_Vital_Dates				
Person_ID	**Person_Name**	**Person_Sex**	**DOB**	**DOD**
3	Richard	Male	1965-08-21 12:00:00 PM	1995-05-11 12:00:00 PM
4	Elizabeth	Female	1967-02-15 12:00:00 PM	
6	Beatrice	Female	1992-10-14 12:00:00 PM	

Query: Select_Generation									
Parent_ID	**Parent_Name**	**Parent_DOB**	**Parent_DOD**		**Parent_Sex**	**Child_ID**	**Child_Name**	**Child_DOB**	**Child_DOD**	**Child_Sex**
3	Richard	1965-08-21 12:00:00 PM	1995-05-11 12:00:00 PM		Male	6	Beatrice	1992-10-14 12:00:00 PM		Female
4	Elizabeth	1967-02-15 12:00:00 PM			Female	6	Beatrice	1992-10-14 12:00:00 PM		Female

Select_Person_Event_History						
State_ID	**Event_ID**	**State_Name**	**Event**	**Influence_Name**	**Observation_Action**	**Event_Timestamp**
3	9	Richard	Birth	Person start: birth	No change	1965-08-21 12:00:00 PM
3	1	Richard	Enrollment	Person, at place: enrollment	No change	1990-03-20 12:00:00 PM
3	1	Richard	Enrollment	Person, enroll: enrollment	Start	1990-03-20 12:00:00 PM
3	1	Richard	Enrollment	Person, residence start: enrollment	No change	1990-03-20 12:00:00 PM
3	4	Richard	Wedding	Person, marry: wedding	No change	1991-04-15 12:00:00 PM
3	*5*	*Richard*	*Birth*	*Person, child born: birth*	*No change*	*1992-10-14 12:00:00 PM*

Figure12: **(continued)**

State_ID	Event_ID	State_Name	Event	Influence_Name	Observation_Action	Event_Timestamp
3	2	Richard	Observation	Person, observed: observation	Stop-Start	1993-11-09 12:00:00 PM
3	6	Richard	Death	Person stop: death	Stop	1995-05-11 12:00:00 PM
3	6	Richard	Death	Person, residence stop: death	No change	1995-05-11 12:00:00 PM
4	10	Elizabeth	Birth	Person start: birth	No change	1967-02-15 12:00:00 PM
4	1	Elizabeth	Enrollment	Person, at place: enrollment	No change	1990-03-20 12:00:00 PM
4	1	Elizabeth	Enrollment	Person, enroll: enrollment	Start	1990-03-20 12:00:00 PM
4	1	Elizabeth	Enrollment	Person, residence start: enrollment	No change	1990-03-20 12:00:00 PM
4	4	Elizabeth	Wedding	Person, marry: wedding	No change	1991-04-15 12:00:00 PM
4	5	Elizabeth	Birth	Person, child born: birth	No change	1992-10-14 12:00:00 PM
4	2	Elizabeth	Observation	Person, observed: observation	Stop-Start	1993-11-09 12:00:00 PM
4	6	Elizabeth	Death	Person, spouse dies: death	No change	1995-05-11 12:00:00 PM
4	7	Elizabeth	Move	Person, move away from: move	No change	1996-08-23 12:00:00 PM
4	7	Elizabeth	Move	Person, move to: move	No change	1996-08-23 12:00:00 PM
4	7	Elizabeth	Move	Person, residence start: move	No change	1996-08-23 12:00:00 PM
4	7	Elizabeth	Move	Person, residence stop: move	No change	1996-08-23 12:00:00 PM
4	3	Elizabeth	Observation	Person, observed: observation	Stop-Start	1997-08-18 12:00:00 PM
6	**5**	**Beatrice**	**Birth**	**Person start: birth**	**Start**	**1992-10-14 12:00:00 PM**
6	**5**	**Beatrice**	**Birth**	**Person, residence start: birth**	**No change**	**1992-10-14 12:00:00 PM**
6	**2**	**Beatrice**	**Observation**	**Person, observed: observation**	**Stop-Start**	**1993-11-09 12:00:00 PM**
6	6	Beatrice	Death	Person, parent dies: death	No change	1995-05-11 12:00:00 PM
6	7	Beatrice	Move	Person, move away from: move	No change	1996-08-23 12:00:00 PM
6	7	Beatrice	Move	Person, move to: move	No change	1996-08-23 12:00:00 PM
6	7	Beatrice	Move	Person, residence start: move	No change	1996-08-23 12:00:00 PM
6	7	Beatrice	Move	Person, residence stop: move	No change	1996-08-23 12:00:00 PM
6	3	Beatrice	Observation	Person, observed: observation	Stop-Start	1997-08-18 12:00:00 PM

Figure 13 illustrates the calculation of demographic rates using the London – New York example (refer to Section A.4 for how to perform these calculations in the example database). The **AD_Time_Intervals** and **AD_Age_Groups** tables contain the time intervals and age groups over which the calculations are to be made and can be edited by the user to define arbitrary new time intervals and age groups. These are used by the queries that calculate person-years and rates to properly split and sum over observed time intervals and age groups. The results of the calculations using the values

displayed in the time intervals and age groups tables are displayed in the **A_Time_Sex_Age_Rates** table. The results are arranged by the time intervals and age groups defined in the **AD_Time_Intervals** and **AD_Age_Groups** tables and further by sex. Counts of births and deaths and total person-years are displayed along with the fertility and mortality rates per 1,000 person-years (never mind the odd enormous rates that result from the very small amount of exposure contained in the example).

A.4.2 "Repeated Rounds" example analytic queries

Results of the analytical queries run on the "repeated rounds" example are displayed below in Figure 14 and Figure 15. These use different time intervals and age groups that are appropriate for this example but are otherwise identical to the analytical queries in the London – New York example (except for the "Select_Vill_HH_Line_Number_Person" query that is obviously specific to this example).

Figure 13:   London – New York example: demographic rates calculations

AD_Time_Intervals

ID	Time_Interval	Start_Date	Stop_Date
1	1990-1994	1990-01-01	1995-01-01
2	1995-1999	1995-01-01	2000-01-01

AD_Age_Groups

ID	Age_Group	Start_Age	Start_Age_Unit	Stop_Age	Stop_Age_Unit
1	0 Years	0	Year	1	Year
2	1-9 Years	1	Year	10	Year
3	10-19 Years	10	Year	20	Year
4	20-39 Years	20	Year	40	Year
5	40+ Years	40	Year	200	Year

Figure 13:   (continued)

			A_Time_Sex_Age_Rates[13]				
Time_Interval	Sex	Age_Group	Births_Observed	Deaths_Observed	Person_Years_Observed	Fx_per_1000	Mx_per_1000
1990-1994	Female	0 Years	0	0	0.9993155373	0	0
1990-1994	Female	1-9 Years	0	0	1.2142368241	0	0
1990-1994	Female	10-19 Years	0	0	0	0	0
1990-1994	Female	20-39 Years	1	0	4.7843942505	209.012875537064	0
1990-1994	Female	40+ Years	0	0	0	0	0
1990-1994	Male	0 Years	0	0	0	0	0
1990-1994	Male	1-9 Years	0	0	0	0	0
1990-1994	Male	10-19 Years	0	0	0	0	0
1990-1994	Male	20-39 Years	0	0	4.7843942505	0	0
1990-1994	Male	40+ Years	0	0	0	0	0
1995-1999	Female	0 Years	0	0	0	0	0
1995-1999	Female	1-9 Years	0	0	2.629705681	0	0
1995-1999	Female	10-19 Years	0	0	0	0	0
1995-1999	Female	20-39 Years	0	0	2.629705681	0	0
1995-1999	Female	40+ Years	0	0	0	0	0
1995-1999	Male	0 Years	0	0	0	0	0
1995-1999	Male	1-9 Years	0	0	0	0	0
1995-1999	Male	10-19 Years	0	0	0	0	0
1995-1999	Male	20-39 Years	0	1	0.3572895277	0	2798.85057487511
1995-1999	Male	40+ Years	0	0	0	0	0

---

[13] For the purpose of calculating person-years events are assumed to occur at 12:00pm (noon), and time intervals are assumed to run from 12:00am (midnight) on the day they start to 12:00am on the day they end. With this definition the time period 2004-01-01 to 2005-01-01 includes every hour of every day of the year 2004 and no amount of time during 2005. Person years are calculated as the number of hours lived between relevant events and beginnings and endings of time intervals divided by 24*365.25.

Figure 14: "Repeated Rounds" example: various analysis queries

	Query: "Select_Person_Names_Sexes_Vital_Dates"			
Person_ID	**Person_Name**	**Person_Sex**	**DOB**	**DOD**
3	V1.HH1.P1	Male	1950-02-12 12:00:00 PM	
4	V1.HH1.P2	Female	1952-11-02 12:00:00 PM	2003-11-02 12:00:00 PM
5	V1.HH1.P3	Male	1976-05-18 12:00:00 PM	
6	V1.HH1.P4	Female	1980-07-25 12:00:00 PM	
8	V1.HH2.P1	Male	1977-01-02 12:00:00 PM	
9	V1.HH2.P2	Female	1981-05-12 12:00:00 PM	
10	V1.HH2.P3	Female	2003-10-19 12:00:00 PM	
13	V2.HH1.P1	Male	1985-08-09 12:00:00 PM	
14	V2.HH1.P2	Female	1986-02-13 12:00:00 PM	
15	V2.HH1.P3	Male	2003-12-06 12:00:00 PM	
17	V2.HH2.P1	Male	1938-04-03 12:00:00 PM	2003-09-12 12:00:00 PM
18	V2.HH2.P2	Female	1940-11-16 12:00:00 PM	2003-06-01 12:00:00 PM
19	V2.HH2.P3	Female	1960-07-17 12:00:00 PM	
20	V2.HH2.P4	Female	1962-11-03 12:00:00 PM	

			Query: "Select_Vill_HH_Line_Number_Person"			
Village Number	**Household Number**	**Line Number**	**Person_Name**	**Person_Sex**	**DOB**	**DOD**
1	1	1	V1.HH1.P1	Male	1950-02-12 12:00:00 PM	
1	1	2	V1.HH1.P2	Female	1952-11-02 12:00:00 PM	2003-11-02 12:00:00 PM
1	1	3	V1.HH1.P3	Male	1976-05-18 12:00:00 PM	
1	1	4	V1.HH1.P4	Female	1980-07-25 12:00:00 PM	
1	2	1	V1.HH2.P1	Male	1977-01-02 12:00:00 PM	
1	2	2	V1.HH2.P2	Female	1981-05-12 12:00:00 PM	
1	2	3	V1.HH2.P3	Female	2003-10-19 12:00:00 PM	
2	1	1	V2.HH1.P1	Male	1985-08-09 12:00:00 PM	
2	1	2	V2.HH1.P2	Female	1986-02-13 12:00:00 PM	
2	1	3	V2.HH1.P3	Male	2003-12-06 12:00:00 PM	
2	2	1	V2.HH2.P1	Male	1938-04-03 12:00:00 PM	2003-09-12 12:00:00 PM
2	2	2	V2.HH2.P2	Female	1940-11-16 12:00:00 PM	2003-06-01 12:00:00 PM
2	2	3	V2.HH2.P3	Female	1960-07-17 12:00:00 PM	
2	2	4	V2.HH2.P4	Female	1962-11-03 12:00:00 PM	

Figure 14:   (continued)

Parent_ID	Parent_Name	Parent_DOB	Parent_DOD	Parent_Sex	Child_ID	Child_Name	Child_DOB	Child_DOD	Child_Sex
			Query: "Select_Generation"						
9	V1.HH2.P2	1981-05-12 12:00:00 PM		Female	10	V1.HH2.P3	2003-10-19 12:00:00 PM		Female
8	V1.HH2.P1	1977-01-02 12:00:00 PM		Male	10	V1.HH2.P3	2003-10-19 12:00:00 PM		Female
14	V2.HH1.P2	1986-02-13 12:00:00 PM		Female	15	V2.HH1.P3	2003-12-06 12:00:00 PM		Male
13	V2.HH1.P1	1985-08-09 12:00:00 PM		Male	15	V2.HH1.P3	2003-12-06 12:00:00 PM		Male

State_ID	Event_ID	State_Name	Event	Influence_Name	Observation_Action	Event_Timestamp
				Select_Person_Event_History		
3	5	V1.HH1.P1	Birth	Birth, starts, person	No change	1950-02-12 12:00:00 PM
3	1	V1.HH1.P1	Initiate	Initiate, enrolls, person	Start	2003-01-05 12:00:00 PM
3	2	V1.HH1.P1	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
3	3	V1.HH1.P1	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
3	4	V1.HH1.P1	Observation	Observation, observes, person	Stop-Start	2003-12-31 12:00:00 PM
4	6	V1.HH1.P2	Birth	Birth, starts, person	No change	1952-11-02 12:00:00 PM
4	1	V1.HH1.P2	Initiate	Initiate, enrolls, person	Start	2003-01-05 12:00:00 PM
4	2	V1.HH1.P2	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
4	3	V1.HH1.P2	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
4	19	V1.HH1.P2	Death	Death, stops, person	Stop	2003-11-02 12:00:00 PM
5	7	V1.HH1.P3	Birth	Birth, starts, person	No change	1976-05-18 12:00:00 PM
5	1	V1.HH1.P3	Initiate	Initiate, enrolls, person	Start	2003-01-05 12:00:00 PM
5	2	V1.HH1.P3	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
5	3	V1.HH1.P3	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
5	4	V1.HH1.P3	Observation	Observation, observes, person	Stop-Start	2003-12-31 12:00:00 PM
6	8	V1.HH1.P4	Birth	Birth, starts, person	No change	1980-07-25 12:00:00 PM
6	1	V1.HH1.P4	Initiate	Initiate, enrolls, person	Start	2003-01-05 12:00:00 PM
6	2	V1.HH1.P4	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
6	3	V1.HH1.P4	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
6	4	V1.HH1.P4	Observation	Observation, observes, person	Stop-Start	2003-12-31 12:00:00 PM
8	**9**	**V1.HH2.P1**	**Birth**	**Birth, starts, person**	**No change**	**1977-01-02 12:00:00 PM**
8	**22**	**V1.HH2.P1**	**Move**	**Move, brings into study area, person**	**Start**	**2003-02-14 12:00:00 PM**
8	**25**	**V1.HH2.P1**	**Move**	**Move, removes from study area, person**	**Stop**	**2003-03-01 12:00:00 PM**
8	**2**	**V1.HH2.P1**	**Observation**	**Observation, observes, person**	**Stop-Start**	**2003-05-05 12:00:00 PM**

Figure 14: **(continued)**

				Select_Person_Event_History		
State_ID	Event_ID	State_Name	Event	Influence_Name	Observation_Action	Event_Timestamp
8	**26**	**V1.HH2.P1**	**Move**	**Move, brings into study area, person**	**Start**	**2003-07-03 12:00:00 PM**
8	**3**	**V1.HH2.P1**	**Observation**	**Observation, observes, person**	**Stop-Start**	**2003-09-02 12:00:00 PM**
8	**11**	**V1.HH2.P1**	**Birth**	**Birth, happens to, person**	**No change**	**2003-10-19 12:00:00 PM**
8	**4**	**V1.HH2.P1**	**Observation**	**Observation, observes, person**	**Stop-Start**	**2003-12-31 12:00:00 PM**
9	10	V1.HH2.P2	Birth	Birth, starts, person	No change	1981-05-12 12:00:00 PM
9	22	V1.HH2.P2	Move	Move, brings into study area, person	Start	2003-02-14 12:00:00 PM
9	2	V1.HH2.P2	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
9	3	V1.HH2.P2	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
9	11	V1.HH2.P2	Birth	Birth, happens to, person	No change	2003-10-19 12:00:00 PM
9	4	V1.HH2.P2	Observation	Observation, observes, person	Stop-Start	2003-12-31 12:00:00 PM
10	11	V1.HH2.P3	Birth	Birth, starts, person	Start	2003-10-19 12:00:00 PM
10	4	V1.HH2.P3	Observation	Observation, observes, person	Stop-Start	2003-12-31 12:00:00 PM
13	12	V2.HH1.P1	Birth	Birth, starts, person	No change	1985-08-09 12:00:00 PM
13	23	V2.HH1.P1	Move	Move, brings into study area, person	Start	2003-03-11 12:00:00 PM
13	2	V2.HH1.P1	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
13	3	V2.HH1.P1	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
13	14	V2.HH1.P1	Birth	Birth, happens to, person	No change	2003-12-06 12:00:00 PM
13	24	V2.HH1.P1	Move	Move, removes from study area, person	Stop	2003-12-20 12:00:00 PM
14	13	V2.HH1.P2	Birth	Birth, starts, person	No change	1986-02-13 12:00:00 PM
14	23	V2.HH1.P2	Move	Move, brings into study area, person	Start	2003-03-11 12:00:00 PM
14	2	V2.HH1.P2	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
14	3	V2.HH1.P2	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
14	14	V2.HH1.P2	Birth	Birth, happens to, person	No change	2003-12-06 12:00:00 PM
14	24	V2.HH1.P2	Move	Move, removes from study area, person	Stop	2003-12-20 12:00:00 PM
15	14	V2.HH1.P3	Birth	Birth, starts, person	Start	2003-12-06 12:00:00 PM
15	24	V2.HH1.P3	Move	Move, removes from study area, person	Stop	2003-12-20 12:00:00 PM
17	15	V2.HH2.P1	Birth	Birth, starts, person	No change	1938-04-03 12:00:00 PM
17	1	V2.HH2.P1	Initiate	Initiate, enrolls, person	Start	2003-01-05 12:00:00 PM
17	2	V2.HH2.P1	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
17	3	V2.HH2.P1	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
17	20	V2.HH2.P1	Death	Death, stops, person	Stop	2003-09-12 12:00:00 PM
18	16	V2.HH2.P2	Birth	Birth, starts, person	No change	1940-11-16 12:00:00 PM
18	1	V2.HH2.P2	Initiate	Initiate, enrolls, person	Start	2003-01-05 12:00:00 PM
18	2	V2.HH2.P2	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM

Figure 14: **(continued)**

State_ID	Event_ID	State_Name	Event	Influence_Name	Observation_Action	Event_Timestamp
				Select_Person_Event_History		
18	21	V2.HH2.P2	Death	Death, stops, person	Stop	2003-06-01 12:00:00 PM
19	17	V2.HH2.P3	Birth	Birth, starts, person	No change	1960-07-17 12:00:00 PM
19	1	V2.HH2.P3	Initiate	Initiate, enrolls, person	Start	2003-01-05 12:00:00 PM
19	2	V2.HH2.P3	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
19	3	V2.HH2.P3	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
19	4	V2.HH2.P3	Observation	Observation, observes, person	Stop-Start	2003-12-31 12:00:00 PM
20	18	V2.HH2.P4	Birth	Birth, starts, person	No change	1962-11-03 12:00:00 PM
20	1	V2.HH2.P4	Initiate	Initiate, enrolls, person	Start	2003-01-05 12:00:00 PM
20	2	V2.HH2.P4	Observation	Observation, observes, person	Stop-Start	2003-05-05 12:00:00 PM
20	3	V2.HH2.P4	Observation	Observation, observes, person	Stop-Start	2003-09-02 12:00:00 PM
20	4	V2.HH2.P4	Observation	Observation, observes, person	Stop-Start	2003-12-31 12:00:00 PM

Figure15: **London – "Repeated Rounds" example: demographic rates calculations**

ID	Time_Interval	Start_Date	Stop_Date
1	2003-Q1	2003-01-01	2003-04-01
2	2003-Q2	2003-04-01	2003-07-01
3	2003-Q3	2003-07-01	2003-10-01
4	2003-Q4	2003-10-01	2004-01-01

ID	Age_Group	Start_Age	Start_Age_Unit	Stop_Age	Stop_Age_Unit
		AD_Age_Groups			
1	0 Years	0	Year	1	Year
2	1-4 Years	1	Year	5	Year
3	5-9 Years	5	Year	10	Year
4	10-14 Years	10	Year	15	Year
5	15-19 Years	15	Year	20	Year
6	20-24 Years	20	Year	25	Year
7	25-29 Years	25	Year	30	Year
8	30-34 Years	30	Year	35	Year

Figure15:   (continued)

ID	Age_Group	Start_Age	Start_Age_Unit	Stop_Age	Stop_Age_Unit
			AD_Age_Groups		
9	35-39 Years	35	Year	40	Year
10	40-44 Years	40	Year	45	Year
11	45-49 Years	45	Year	50	Year
12	50-54 Years	50	Year	55	Year
13	55-59 Years	55	Year	60	Year
14	60-64 Years	60	Year	65	Year
15	65-69 Years	65	Year	70	Year
16	70-74 Years	70	Year	75	Year
17	75-79 Years	75	Year	80	Year
18	80-84 Years	80	Year	85	Year
19	85-89 Years	85	Year	90	Year
20	90-94 Years	90	Year	95	Year
21	95-99 Years	95	Year	100	Year
22	100+ Years	100	Year	200	Year

**Figure15:   London – "Repeated Rounds" example:
demographic rates calculations**

Time_Interval	Sex	Age_Group	Births_Observed	Deaths_Observed	Person_Years_Observed	Fx_per_1000	Mx_per_1000
			A_Time_Sex_Age_Rates				
2003-Q1	Female	0 Years	0	0	0	0	0
2003-Q1	Female	1-4 Years	0	0	0	0	0
2003-Q1	Female	5-9 Years	0	0	0	0	0
2003-Q1	Female	10-14 Years	0	0	0	0	0
2003-Q1	Female	15-19 Years	0	0	0.0561259411	0	0
2003-Q1	Female	20-24 Years	0	0	0.3586584531	0	0
2003-Q1	Female	25-29 Years	0	0	0	0	0
2003-Q1	Female	30-34 Years	0	0	0	0	0
2003-Q1	Female	35-39 Years	0	0	0	0	0
2003-Q1	Female	40-44 Years	0	0	0.4681724846	0	0
2003-Q1	Female	45-49 Years	0	0	0	0	0

Figure15: **(continued)**

			A_Time_Sex_Age_Rates				
Time_Interval	**Sex**	**Age_Group**	**Births_Observed**	**Deaths_Observed**	**Person_Years_Observed**	**Fx_per_1000**	**Mx_per_1000**
2003-Q1	Female	50-54 Years	0	0	0.2340862423	0	0
2003-Q1	Female	55-59 Years	0	0	0	0	0
2003-Q1	Female	60-64 Years	0	0	0.2340862423	0	0
2003-Q1	Female	65-69 Years	0	0	0	0	0
2003-Q1	Female	70-74 Years	0	0	0	0	0
2003-Q1	Female	75-79 Years	0	0	0	0	0
2003-Q1	Female	80-84 Years	0	0	0	0	0
2003-Q1	Female	85-89 Years	0	0	0	0	0
2003-Q1	Female	90-94 Years	0	0	0	0	0
2003-Q1	Female	95-99 Years	0	0	0	0	0
2003-Q1	Female	100+ Years	0	0	0	0	0
2003-Q1	Male	0 Years	0	0	0	0	0
2003-Q1	Male	1-4 Years	0	0	0	0	0
2003-Q1	Male	5-9 Years	0	0	0	0	0
2003-Q1	Male	10-14 Years	0	0	0	0	0
2003-Q1	Male	15-19 Years	0	0	0.0561259411	0	0
2003-Q1	Male	20-24 Years	0	0	0	0	0
2003-Q1	Male	25-29 Years	0	0	0.2751540041	0	0
2003-Q1	Male	30-34 Years	0	0	0	0	0
2003-Q1	Male	35-39 Years	0	0	0	0	0
2003-Q1	Male	40-44 Years	0	0	0	0	0
2003-Q1	Male	45-49 Years	0	0	0	0	0
2003-Q1	Male	50-54 Years	0	0	0.2340862423	0	0
2003-Q1	Male	55-59 Years	0	0	0	0	0
2003-Q1	Male	60-64 Years	0	0	0.2340862423	0	0
2003-Q1	Male	65-69 Years	0	0	0	0	0
2003-Q1	Male	70-74 Years	0	0	0	0	0
2003-Q1	Male	75-79 Years	0	0	0	0	0
2003-Q1	Male	80-84 Years	0	0	0	0	0
2003-Q1	Male	85-89 Years	0	0	0	0	0
2003-Q1	Male	90-94 Years	0	0	0	0	0
2003-Q1	Male	95-99 Years	0	0	0	0	0
2003-Q1	Male	100+ Years	0	0	0	0	0
2003-Q2	Female	0 Years	0	0	0	0	0

Figure15:   (continued)

Time_Interval	Sex	Age_Group	Births_Observed	Deaths_Observed	Person_Years_Observed	Fx_per_1000	Mx_per_1000
				A_Time_Sex_Age_Rates			
2003-Q2	Female	1-4 Years	0	0	0	0	0
2003-Q2	Female	5-9 Years	0	0	0	0	0
2003-Q2	Female	10-14 Years	0	0	0	0	0
2003-Q2	Female	15-19 Years	0	0	0.2491444216	0	0
2003-Q2	Female	20-24 Years	0	0	0.4982888433	0	0
2003-Q2	Female	25-29 Years	0	0	0	0	0
2003-Q2	Female	30-34 Years	0	0	0	0	0
2003-Q2	Female	35-39 Years	0	0	0	0	0
2003-Q2	Female	40-44 Years	0	0	0.4982888433	0	0
2003-Q2	Female	45-49 Years	0	0	0	0	0
2003-Q2	Female	50-54 Years	0	0	0.2491444216	0	0
2003-Q2	Female	55-59 Years	0	0	0	0	0
2003-Q2	Female	60-64 Years	0	1	0.1683778234	0	5939.0243905481
2003-Q2	Female	65-69 Years	0	0	0	0	0
2003-Q2	Female	70-74 Years	0	0	0	0	0
2003-Q2	Female	75-79 Years	0	0	0	0	0
2003-Q2	Female	80-84 Years	0	0	0	0	0
2003-Q2	Female	85-89 Years	0	0	0	0	0
2003-Q2	Female	90-94 Years	0	0	0	0	0
2003-Q2	Female	95-99 Years	0	0	0	0	0
2003-Q2	Female	100+ Years	0	0	0	0	0
2003-Q2	Male	0 Years	0	0	0	0	0
2003-Q2	Male	1-4 Years	0	0	0	0	0
2003-Q2	Male	5-9 Years	0	0	0	0	0
2003-Q2	Male	10-14 Years	0	0	0	0	0
2003-Q2	Male	15-19 Years	0	0	0.2491444216	0	0
2003-Q2	Male	20-24 Years	0	0	0	0	0
2003-Q2	Male	25-29 Years	0	0	0.2491444216	0	0
2003-Q2	Male	30-34 Years	0	0	0	0	0
2003-Q2	Male	35-39 Years	0	0	0	0	0
2003-Q2	Male	40-44 Years	0	0	0	0	0
2003-Q2	Male	45-49 Years	0	0	0	0	0
2003-Q2	Male	50-54 Years	0	0	0.2491444216	0	0
2003-Q2	Male	55-59 Years	0	0	0	0	0

Figure15: (continued)

			A_Time_Sex_Age_Rates				
Time_Interval	Sex	Age_Group	Births_Observed	Deaths_Observed	Person_Years_Observed	Fx_per_1000	Mx_per_1000
2003-Q2	Male	60-64 Years	0	0	0.006844627	0	0
2003-Q2	Male	65-69 Years	0	0	0.2422997947	0	0
2003-Q2	Male	70-74 Years	0	0	0	0	0
2003-Q2	Male	75-79 Years	0	0	0	0	0
2003-Q2	Male	80-84 Years	0	0	0	0	0
2003-Q2	Male	85-89 Years	0	0	0	0	0
2003-Q2	Male	90-94 Years	0	0	0	0	0
2003-Q2	Male	95-99 Years	0	0	0	0	0
2003-Q2	Male	100+ Years	0	0	0	0	0
2003-Q3	Female	0 Years	0	0	0	0	0
2003-Q3	Female	1-4 Years	0	0	0	0	0
2003-Q3	Female	5-9 Years	0	0	0	0	0
2003-Q3	Female	10-14 Years	0	0	0	0	0
2003-Q3	Female	15-19 Years	0	0	0.2518822724	0	0
2003-Q3	Female	20-24 Years	0	0	0.5037645448	0	0
2003-Q3	Female	25-29 Years	0	0	0	0	0
2003-Q3	Female	30-34 Years	0	0	0	0	0
2003-Q3	Female	35-39 Years	0	0	0	0	0
2003-Q3	Female	40-44 Years	0	0	0.5037645448	0	0
2003-Q3	Female	45-49 Years	0	0	0	0	0
2003-Q3	Female	50-54 Years	0	0	0.2518822724	0	0
2003-Q3	Female	55-59 Years	0	0	0	0	0
2003-Q3	Female	60-64 Years	0	0	0	0	0
2003-Q3	Female	65-69 Years	0	0	0	0	0
2003-Q3	Female	70-74 Years	0	0	0	0	0
2003-Q3	Female	75-79 Years	0	0	0	0	0
2003-Q3	Female	80-84 Years	0	0	0	0	0
2003-Q3	Female	85-89 Years	0	0	0	0	0
2003-Q3	Female	90-94 Years	0	0	0	0	0
2003-Q3	Female	95-99 Years	0	0	0	0	0
2003-Q3	Female	100+ Years	0	0	0	0	0
2003-Q3	Male	0 Years	0	0	0	0	0
2003-Q3	Male	1-4 Years	0	0	0	0	0

Figure15:   (continued)

				A_Time_Sex_Age_Rates			
Time_Interval	Sex	Age_Group	Births_Observed	Deaths_Observed	Person_Years_Observed	Fx_per_1000	Mx_per_1000
2003-Q3	Male	5-9 Years	0	0	0	0	0
2003-Q3	Male	10-14 Years	0	0	0	0	0
2003-Q3	Male	15-19 Years	0	0	0.2518822724	0	0
2003-Q3	Male	20-24 Years	0	0	0	0	0
2003-Q3	Male	25-29 Years	0	0	0.4969199179	0	0
2003-Q3	Male	30-34 Years	0	0	0	0	0
2003-Q3	Male	35-39 Years	0	0	0	0	0
2003-Q3	Male	40-44 Years	0	0	0	0	0
2003-Q3	Male	45-49 Years	0	0	0	0	0
2003-Q3	Male	50-54 Years	0	0	0.2518822724	0	0
2003-Q3	Male	55-59 Years	0	0	0	0	0
2003-Q3	Male	60-64 Years	0	0	0	0	0
2003-Q3	Male	65-69 Years	0	1	0.2012320329	0	4969.38775397125
2003-Q3	Male	70-74 Years	0	0	0	0	0
2003-Q3	Male	75-79 Years	0	0	0	0	0
2003-Q3	Male	80-84 Years	0	0	0	0	0
2003-Q3	Male	85-89 Years	0	0	0	0	0
2003-Q3	Male	90-94 Years	0	0	0	0	0
2003-Q3	Male	95-99 Years	0	0	0	0	0
2003-Q3	Male	100+ Years	0	0	0	0	0
2003-Q4	Female	0 Years	0	0	0.1998631075	0	0
2003-Q4	Female	1-4 Years	0	0	0	0	0
2003-Q4	Female	5-9 Years	0	0	0	0	0
2003-Q4	Female	10-14 Years	0	0	0	0	0
2003-Q4	Female	15-19 Years	1	0	0.2203969884	4537.26708000698	0
2003-Q4	Female	20-24 Years	1	0	0.501026694	1995.90163952422	0
2003-Q4	Female	25-29 Years	0	0	0	0	0
2003-Q4	Female	30-34 Years	0	0	0	0	0
2003-Q4	Female	35-39 Years	0	0	0	0	0
2003-Q4	Female	40-44 Years	0	0	0.501026694	0	0
2003-Q4	Female	45-49 Years	0	0	0	0	0
2003-Q4	Female	50-54 Years	0	1	0.0889801506	0	11238.461536162
2003-Q4	Female	55-59 Years	0	0	0	0	0
2003-Q4	Female	60-64 Years	0	0	0	0	0

Figure15: **(continued)**

			A_Time_Sex_Age_Rates				
Time_Interval	Sex	Age_Group	Births_Observed	Deaths_Observed	Person_Years_Observed	Fx_per_1000	Mx_per_1000
2003-Q4	Female	65-69 Years	0	0	0	0	0
2003-Q4	Female	70-74 Years	0	0	0	0	0
2003-Q4	Female	75-79 Years	0	0	0	0	0
2003-Q4	Female	80-84 Years	0	0	0	0	0
2003-Q4	Female	85-89 Years	0	0	0	0	0
2003-Q4	Female	90-94 Years	0	0	0	0	0
2003-Q4	Female	95-99 Years	0	0	0	0	0
2003-Q4	Female	100+ Years	0	0	0	0	0
2003-Q4	Male	0 Years	0	0	0.038329911	0	0
2003-Q4	Male	1-4 Years	0	0	0	0	0
2003-Q4	Male	5-9 Years	0	0	0	0	0
2003-Q4	Male	10-14 Years	0	0	0	0	0
2003-Q4	Male	15-19 Years	0	0	0.2203969884	0	0
2003-Q4	Male	20-24 Years	0	0	0	0	0
2003-Q4	Male	25-29 Years	0	0	0.501026694	0	0
2003-Q4	Male	30-34 Years	0	0	0	0	0
2003-Q4	Male	35-39 Years	0	0	0	0	0
2003-Q4	Male	40-44 Years	0	0	0	0	0
2003-Q4	Male	45-49 Years	0	0	0	0	0
2003-Q4	Male	50-54 Years	0	0	0.250513347	0	0
2003-Q4	Male	55-59 Years	0	0	0	0	0
2003-Q4	Male	60-64 Years	0	0	0	0	0
2003-Q4	Male	65-69 Years	0	0	0	0	0
2003-Q4	Male	70-74 Years	0	0	0	0	0
2003-Q4	Male	75-79 Years	0	0	0	0	0
2003-Q4	Male	80-84 Years	0	0	0	0	0
2003-Q4	Male	85-89 Years	0	0	0	0	0
2003-Q4	Male	90-94 Years	0	0	0	0	0
2003-Q4	Male	95-99 Years	0	0	0	0	0
2003-Q4	Male	100+ Years	0	0	0	0	0

In examining these analytical results it is critical to note that the databases and queries (SQL code) used in both examples are *identical*. The only differences between

the London – New York and "repeated rounds" examples are that they contain different metadata, different primary data and use different time intervals and age groups.

A.5 Definitions of temporal elements

This section is adapted from Benzler and Clark (2005).

A.5.1 Time, measures of time, and valid time

Time is universal, one-dimensional, dense and unbounded. A single time domain exists at all locations, leading to a general notion of concurrence (*universal*). Individual elements constituting the time domain have no extent within the domain (zero duration) and are unambiguously identified by ordered, unique values of a single, numeric attribute called "position" (*one-dimensional*). Between any two elements it is possible to insert an additional element (*dense*), and given this it is always possible to insert a new element *before* the first and *after* the last element (*unbounded*).

A.5.1.1 Measures of time

There are five fundamental measures of time. These allow us to conceptualize and manipulate the time domain irrespective of the meaning that may be associated with time.

A.5.1.1.1 Time element

A time element is the basic element constituting the time domain. A time element can be located at any position within the time domain and has zero duration.

A.5.1.1.2 Time point

A time point identifies a single element in the time domain with a position. Because it simply identifies a time element, a time point also has zero duration. A time point is labeled with a single numeric value.

A.5.1.1.3 Time duration

A time duration identifies an extent within the time domain. A time duration does not have a position within the time domain, nor is a time duration a single set of time elements. A time duration is labeled with a single numeric value denoting its extent in the time domain. Because the elements of the time domain are ordered, a time duration can have directionality and can extend in either direction along the time domain.

A.5.1.1.4 Time interval

A time interval identifies both an extent and a position within the time domain. A time interval consists of a bounded, infinite set of time elements within the time domain. A time interval is labeled with either: 1) two single numeric values corresponding to the positions of its first and last time elements, or 2) a single numeric value corresponding to its first (or last) time element and a single numeric value corresponding to its duration (extent in the time domain, including direction).

A.5.1.1.5 Time set

A time set identifies a finite set of non-overlapping time intervals within the time domain.

A.5.1.2 Measures of time with meaning: valid time

Here, *meaning* refers to the salience of a fact; something that is both perceived and important in the scope of human experience. Combining this notion of meaning with measures of time yields Valid Time (VT). More precisely VT is the time when a fact is true in the real world (Jensen et al., 1998). VT associates a true proposition with a measure of time, and consequently, VT also takes five basic forms.

A.5.1.2.1 Instant

An instant is the association of a fact with a time element. The meaning of an instant is well-defined while its position in the time domain is not. Like a time point, an instant has zero duration.

A.5.1.2.2 Event

An event is the association of a fact with a time point. Both the meaning of an event and its position in the time domain are well-defined.

A.5.1.2.3 Period

A time period is the association of a fact with a time duration. The proposition that remains true throughout a period is well-defined, while the position of the period within the time domain is not.

A.5.1.2.4 State

A state is the association of a fact with a time interval. The proposition that remains true throughout the state, the duration of the state and the position of the state within the time domain are all well-defined.

A.5.1.2.5 Pattern

A time pattern is the association of a fact with a time set. Both the fact that is true during the time intervals that constitute the pattern, and the extent and position of those time intervals within the time domain are well-defined.

A.5.2 Measures of time in practice, and precision

In practice we use both the raw measures of time and the measures of VT. For example, the raw measures of time are manipulated to construct various calendars and the systems used to convert between them and are also used to construct and calibrate clocks.

However, it is the VT measures that are most common in our daily lives and most likely to be stored and manipulated in a database. We routinely refer to events that affect us; births, deaths, marriages, divorces, the start of the work day, or the end of the month. Likewise states are a natural part of our everyday vernacular; Jack and Jill's marriage, the life of Mozart, or the Second World War. The other VT measures are perhaps less obvious, but they are also part of our daily lives. A state of having flu that

you had last year *sometime* is a duration, although in this case the uncertainty about its position in the time domain is bounded. A woman's pregnancies (pregnancy states) compose a pattern, and instants often describe marginally salient events whose position in the time domain is not well-defined, such as the purchase of a lava lamp *sometime* in the past.

In reality the measurement of time is a messy business and the theoretical precision assumed in the preceding sections is never possible. Instead every fact is mapped to the time domain with some degree of fuzziness. For example we cannot pinpoint the precise, zero-duration time element associated with a birth, but we can put bounds on when the birth took place. Likewise all events are recorded with some degree of precision that is never perfect but often knowable. In order to preserve and store the maximum amount of information and to exclude any implicit assumption about precision, a new type of timestamp is necessary that effectively stores all time points as intervals that correspond to the degree of imprecision associated with the measurement of the time point. A full discussion of temporal measures with explicit precision is presented elsewhere (Benzler and Clark, 2005).

The foregoing observation is of significant practical importance when recording the histories of human beings in the developing world where dates are often not known with great precision. It is not uncommon to have a respondent provide the date of a birth, death, wedding or some other event as "sometime in month x of year y", or just "sometime in year y", or even more challenging "in the rainy season several years ago"!