# A unified framework of demographic time

Tim Riffe*[1], Jonas Schöley[2,3], and Francisco Villavicencio[2,3]

[1]Max Planck Institute for Demographic Research
[2]University of Southern Denmark
[3]Max-Planck Odense Center on the Biodemography of Aging

April 20, 2017

*riffe@demogr.mpg.de

**Abstract**

Demographic thought and practice is largely conditioned by the Lexis diagram, a two-dimensional graphical representation of the identity between age, period, and birth cohort. This relationship does not account for remaining years of life, total length of life, or time of death, whose use in demographic research is both underrepresented and incompletely situated. We describe an identity between these six demographic time measures, and generalize this relationship to time measures derived from an arbitrary number of events in calendar time. We describe the sub-identities that pertain to the six-way demographic time identity, and provide a topological overview of the diagrams that pertain to these identities in both two and three dimensions. We provide an application of this framework on the measurement of late-life disability prevalence.

**Keywords.** Age structure, formal demography, data visualization, age period cohort.

# 1 Introduction

In the course of training, all demographers are introduced to the Lexis diagram, a convenient graphical identity between the three main time measures used to structure demographic stocks and flows: Age, period, and birth cohort. This representation does not account for time of death, time until death, or length of life, which may be of interest to researchers ~~and policy makers~~ as structuring rather than random variables in order to capture variation in demographic data. ~~This popular representation does not account for remaining years of life and other related time indices that may be of interest to researchers and policy makers.~~

We wish to draw attention to three time indices that are complementary to age (A), period (P) and birth cohort (C). The first such index is time-to-death, which we refer to as "thanatological age" (T) in contrast to "chronological age" (A). The second index is death cohort (D), which groups all individuals (of different ages) dying in the same time period. Finally, lifespan (L) or individual age-at-death itself is an index by which data may be structured. We therefore have six time measures in total to relate. We call these measures of demographic time because each, except period, depends on the timing of birth, death, or both.

The Lexis diagram can be understood as an APC plane that relates age, period, and birth cohort. Other such planes are also identifiable. The "thanatological" "dual" of APC is an identity between thanatological age, period, and death cohort, TPD. A third identity relates thanatological age, chronological age, and lifespan, TAL. A fourth identity relates lifespan, birth cohort, and death cohort, LCD. We call three-way identities of this sort "triad identities".

Each of these four triad identities (APC, TPD, TAL, and LCD) is sufficiently described by any two of its constituent indices~~, making the third index redundant~~. For instance, if the exact age of an individual at a particular time is known, the birth cohort to which he or she belongs can be immediately derived. Each of these four identities also lacks a major dimension of time. The TAL identity lacks calendar time, the LCD identity is ageless, APC lacks an endpoint in time, and TPD lacks a starting point in time. To our knowledge, the only triad identity that has received serious treatment at the time of this writing is the APC identity. Different aspects of the APC identity have been discussed since at least 1868 (Knapp 1868), and discussion remains lively today. Here we relate the six major indices of time in a geometric identity, in much the same spirit as the work on APC relationships done between the late 1860s and mid 1880s.[1]

Our goal is to describe the geometric identity between all six measures of demographic time, a hexad identity, that may be useful or an intuitive referent for demographers in the same way as the Lexis diagram. At the same time, this identity relates the four triad identities we have mentioned. We give a bottom-up description of how such temporal identities can be derived from an arbitrary number of events in calendar time in a simple and general event-duration relationship. These more general event-duration foundations facilitate comparison of the demographic time framework with other temporal designs found in the literature, such as the higher-order temporal model of Brinks et al. (2014), or the marriage cohort hexad identity described by Lexis (1875). The framework we describe is general and adaptable for such event history scenarios. We begin by defining some terms used throughout the manuscript. We then explore all combinations of two time measures, the dyadic relationships, followed by the four triad identities, and finally the hexad identity. We give a systematic topological overview of the different elements of demographic time.

Just as the Lexis diagram has been a fundamental instrument to teach demography for decades, we hope that the demographic time measures and their graphical depictions presented here will be helpful to teachers and young demographers. The temporal relationships we describe will also be useful for researchers to better detect and understand patterns data, and for methodologists to better account for the structure of data in demographic methods or statistical designs.

# 2 Definitions

## 2.1 Technical terminology

In describing this relationship we attempt to adhere to a rigorous terminology. The following list describes some of the more important terms we use.

**Demographic time measures** are any of the six time indices discussed to describe demographic time: chronological age (A), period (P), birth cohort (C), thanatological age (T), lifespan (L), and death cohort (D).

**Dyads, triads, and hexads** are any set of two, three, or six unique time measures, respectively.

**A triad identity** is a triad with the property that each of its members can be derived from the other two with no additional information. There are four triad identities: APC, TPD, TAL, and LCD.

**A temporal plane** is any $(x, y)$-mapping of a dyad of time measures.

Using this terminology, we say that the "Lexis" measures constitute a triad identity between chronological age, period, and birth cohort. Each dyad combination of elements in this identity can be mapped to a temporal plane, the Lexis diagram. If we know that Mindel turned 50 on the 21st of May, 1963, then we also can derive that she was born on the 21st of May, 1913. Hence, any two pieces of information in this case will give the third, and the same holds for the other triad identities.

[1]See e.g., Keiding (2011) for an overview of that literature.

## 2.2 Time measures

We describe time in terms of years, the dominant time scale for human demography, although all relationships are scalable to any time unit. We therefore speak of calendar time. We also describe the framework in terms of human lifespans, although it applies in a more general sense to any durations observed over time. This is to say, birth may be translated to entry, and death to exit, or any other absorbing state. The six measures of time we consider are defined in Table 1, both in the demographic sense we describe, as well as in a more general event history interpretation.

Table 1: Definitions of the six time measures.

| **Time measure** | **Short** | **Demographic def.** | **Event history def.** |
|---|---|---|---|
| chronological age | A | Time since birth | Time since start of exposure |
| period | P | calendar time | calendar time |
| birth cohort | C | calendar time of birth | calendar time of exposure start |
| thanatological age | T | time until death | time until event |
| death cohort | D | calendar time of death | calendar time of event |
| lifespan | L | duration of life | duration of exposure |

The concepts of thanatological age and death cohorts are likely less familiar to readers than the other measures we consider. Thanatological age invokes the remaining lifetime until death, the information approximated with life expectancy. This term is sometimes referred to in the literature as life left, time to death, remaining lifespan, follow-up duration, residual life, or reverse time. Chronological and thanatological age are in this way complementary, duals. On the other hand, cohorts in general associate individuals that share a characteristic. In demography the grouping characteristic is often a combination of place and time, such as a cohort of young demographers passing through a particular graduate program. These concepts are analogous to the ideas of birth and death cohorts we use here. The deaths of a given year are not usually referred to as a death cohort, although this concept was already introduced by Brouard (1986) as “génération de décès” in a retrospective study of the French population during the twentieth century. In the time preceding death, the members of a given death cohort have much in common, despite heterogeneity with respect to time of birth. If the reader accepts this premise, then the abstract construct of a death cohort is also meaningful in the way that other cohort measures are. In event history or non-human contexts, anologs to death cohorts in this framework may be even more meaningful.

Much of the work of demography is directed at the study of lifespan. Lifespan is synonymous both with longevity, chronological age at death, and thanatological age at birth. One’s ultimate completed lifespan is constant throughout life, though we have no knowledge of it until death: It is assigned retrospectively. Demographers have more often used lifespan or age-at-death as a measure of mortality, or similar, than as a measure on which to compare individuals or structure data.

Treating lifespan, death cohorts, and thanatological age as temporal structuring variables enables new classes of comparisons, models of understanding, and discovery, akin to those unlocked by breaking down demographic phenomena by chronological age, period, and birth cohort. The following sections, in this sense, provide an exhaustive classification of the ways in which these six measures of time can be juxtaposed to such ends.

# 3 From dyads to the triad identities

We distinguish between two kinds of dyads: informative dyads and uninformative dyads. Informative dyads are any pair of measures from which a third time measure can be derived, forming a triad identity. There are $15 = \binom{6}{2}$ possible dyads in our set of time measures, 12 of which are informative, and three of which have no derived time measure, and are therefore called uninformative. For instance, if we take the dyad TA, L is the derived measure, and TAL the corresponding triad identity. In contrast, nothing can be derived from the LP dyad: You can have an eventual lifespan of 100 in the year 2016 and still be alive with the same eventual lifespan in 2017.

In this section we systematically map each dyad to its temporal plane, and we synthesize these into the four primary identities and their essential diagrams. We first discuss the choice between mapping dyads to Cartesian coordinates or to isotropic coordinates, which constrain the scales of all measures to be equal. We then systematically render the 15 dyad-based diagrams that can be derived from the six time measures. Of these 15, 12 diagrams can be distilled into just four, the triad identity diagrams. Each triad identity diagram is then briefly discussed with suggested or speculated applications.

## 3.1 The question of mapping

Any mapping of two time measures to an $(x, y)$ coordinate system constitutes a temporal plane. If the two given time measures are members of the same triad identity, the third member is a derived measure. If we assign A to $y$ and P to $x$, thereby implying C (and the APC triad identity) we state this relationship explicitly by writing AP(C). The temporal plane that corresponds to this informative dyad is the contemporary representation of the Lexis diagram (Lexis 1875, Pressat 1961). The informative dyads AC(P) and CP(A) also belong to the Lexis identity, but imply different less-common rotations and projections of the Lexis diagram.[2]

For each dyad there is a fundamental question of how to map the constituent coordinates to a Cartesian temporal plane. Typically one forces parity between time units within a specified dyad, mapping one element directly to $x$ and the second element directly to $y$, resulting in a 90° angle between the $x$ and $y$ axes. In this case it is conventional to force a unity aspect ratio between the $x$ and $y$ axes, such that the derived measure, if any, is then *accidentally* present in a 45° ascending or descending angle, depending on the dyad and axis orientation.

It has long been noted (Lexis 1875, Perozzo 1880) that the derived time measure (usually birth cohort) is longer than either the age or period axes when plotted at 45°. If a right angle and unity aspect ratio is forced between the dyad, the derived measure is always stretched by $\sqrt{2}$, or 41%. In the case of informative dyads, another logical mapping is to translate to $(x, y)$ coordinates that force 60° angles between the three measures. Such a mapping ensures that the spatial units are equal for the three measures, and we therefore refer to it as the isotropic mapping. The isotropic mapping is comparable to using ternary or barycentric coordinate systems. The primary justification for isotropic demographic

[2] While uncommon as diagram orientations, these latter two dyads are used to tabulate data for different kinds of rates and probabilities. Measures based on the AC dydad are variously referred to as Type III rates (Caselli et al. 2006) or vertical parallelograms (Wilmoth et al. 2007). The CP dyad delineates Type II rates (Caselli et al. 2006), also known as horizontal parallelograms, for instance used to calculate cohort lifetables in the Human Mortality Database (Wilmoth et al. 2007).

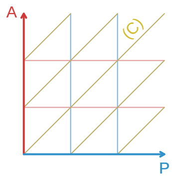
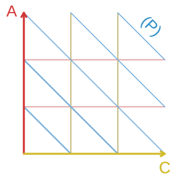
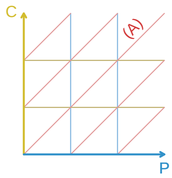
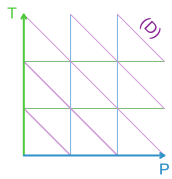
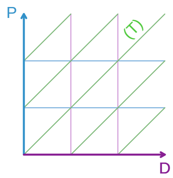
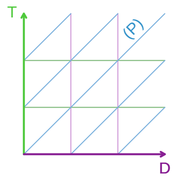
surfaces comes from a data visualization perspective, where it may be hypothesized that the viewer's ability to compare slopes is hindered if time coordinates are not on the same scale. Under the isotropic representation, the three variants of each triad identity are simple rotations of one another, and they require no rescaling. In this paper, all two-dimensional diagrams are rendered in Cartesian rather than isotropic coordinates.

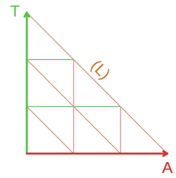
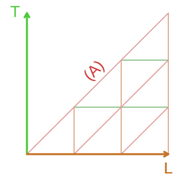
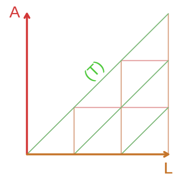
## 3.2 Dyads to diagrams

Each of the 15 dyads, an explanation or simple example, and the corresponding diagram representations are summarized in Table 2. Each informative dyad is a subset consisting of two elements from one of the four triad identities (APC, TPD, TAL, LCD), which we analyze in detail in further sections. The uninformative dyads are simply pairs of time measures that do not have a derived measure, and therefore are not contained in any of these four triad identities.

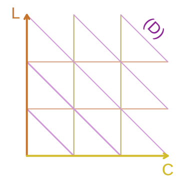
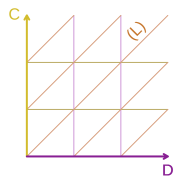
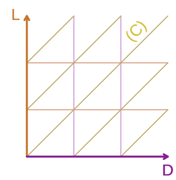
Table 2: All dyadic juxtapositions of the six measures of demographic time.

*Note:* The temporal planes are named after the two given time scales. The derived scale is appended in parentheses. Contrary to mathematical convention we name the ordinate scale first and the abscissa scale second. This is to be consistent with the established *APC* and *ACP* terms.

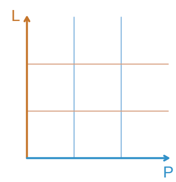
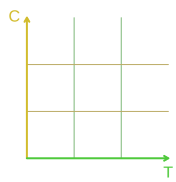
| Relationship | Description | Cartesian |
|---|---|---|
| | **Variants of APC** | |
| $AP(C)$<br>$C = P - A$ | The AP(C) temporal plane constitutes the classical Lexis diagram. |  |
| $AC(P)$<br>$P = C + A$ | The AC(P) temporal plane is equivalent to the Lexis diagram except birth cohort is given and period is derived rather than the other way around. |  |
| $CP(A)$<br>$A = P - C$ | The CP(A) temporal plane is equivalent to the Lexis diagram except birth cohorts are given and age is derived rather than the other way around. |  |
| | **Variants of TPD** | |
| $TP(D)$<br>$D = P + T$ | Helen had 30 years of life left (T) in 1971 (P) and therefore belonged to the 2001 death cohort (D) |  |
| $PD(T)$<br>$T = D - P$ | Mindel died in 1973 (D). In 1953 (P) she had 20 years left to live (T). |  |
| $TD(P)$<br>$P = D - T$ | Irene died in 1974 (D). When she had 30 remaining years of life (T) the year must have been 1944 (P). |  |
| | **Variants of TAL** | |

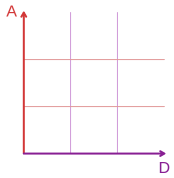
| | | |
|---|---|---|
| $TA(L)$<br>$L = T + A$ | The time already lived and the time still left sum up to the total lifespan. |  |
| $TL(A)$<br>$A = L - T$ | Helen lived to the age of 86 (L). When she had 20 years left (T) she must have been 66 (A). |  |
| $AL(T)$<br>$T = A - L$ | Tim is 34 years old (A) and will live to the age of 96 (L), leaving him 62 years (T) to settle affairs. |  |

## Variants of LCD

| | | |
|---|---|---|
| $LC(D)$<br>$D = C + L$ | Àngels was born in 1940 (C) and she lived to be 64 (L), implying an untimely death in 2004 (D) |  |
| $CD(L)$<br>$L = D - C$ | Pascal was born in 1893 (C) and died in 1964 (D), implying a lifespan of 71 (L), or so. |  |
| $LD(C)$<br>$C = D - L$ | Margaret died in Dec., 1995 (D) with a completed lifespan of 96 (L), putting her birth year in 1900 (C). |  |

## The Uninformative Dyads

| | | |
|---|---|---|
| LP(-) | The LP plane is *non-informative*. No additional measures can be derived knowing just lifespan and period. |  |
| CT(-) | The CT plane is *non-informative*. No additional measures can be derived knowing just birth cohort and thanatological age. |  |

| | | |
|---|---|---|
| AD(-) | The AD plane is *non-informative*. No additional measures can be derived knowing just death cohort and age. |   |

Most of what we know about how rates change over age and time comes from the very first juxtaposition in Table 2, AP(C). While CP(A) and AC(P) are statistically redundant when exact times are used, they are not fully redundant if based on discrete double-classification of data, as often provided in aggregated official statistics. Double classified data are found on the APC diagram in the shape of squares (AP), horizontal parallelograms (AC) and vertical paralellograms (PC), and these are commonly used to compute different kinds of demographic rates and probabilities (Caselli et al. 2006, p. 63). The other dyadic juxtapositions (involving the measures T, D, or L) can be considered as either rare or novel ways of structuring or viewing temporal variation in demography, and these imply new families of rates and probabilities.

## 3.3 The triad identities

There are $20 = \binom{6}{3}$ ways to choose three time indices out of six, of which four form a triad identity: APC, TPD, TAL, and LCD. Given the three time measures from any of the triad identities, one can derive no further time measures. If one selects three random time indices that do not form any of these four triad identities ($20 - 4 = 16$ possibilities), this property does not hold. For instance, in the triad APT, age and period are not sufficient to determine thanatological age. Given the triad APT one can however derive the remaining three time measures.

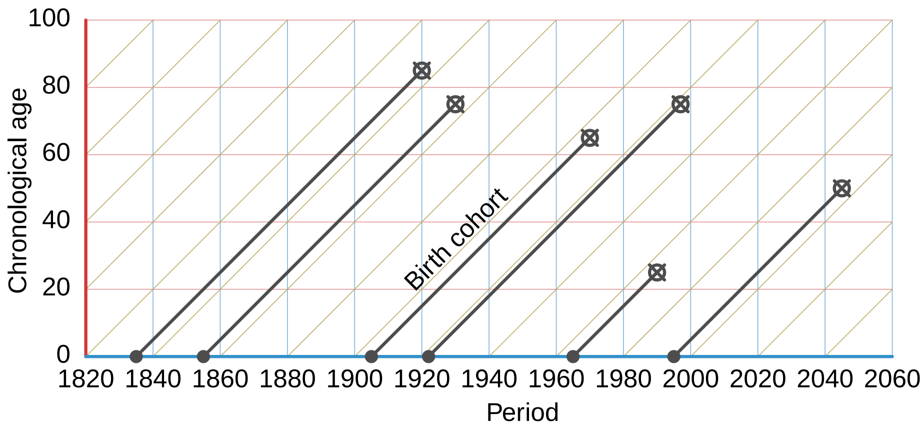
Triad identities are more meaningful than uninformative dyads. This is so even in the absence of data, due to the underlying relationship between measures. Each of the triad identities can accommodate some version of a lifeline, for instance. In the following, we therefore lay out the four primary diagrams that belong to the triad identities. The question of which diagram mapping is relevant to a given demographic phenomena is a function of patterns in the data. The best diagram is the one that captures all meaningful variation in the data. If APC highlights meaningful variation in a phenomenon, then its representation as such is useful. The same holds for the other identities. With each diagram in following we comment or speculate on its potential uses.

### 3.3.1 APC: Chronological age, period, and birth cohort

The so-called Lexis diagram has long been used in demography as a conceptual tool for structuring data, observations, and rate estimation, as inspiration for work on statistical identification, and as the coordinate basis of contemporary Lexis-surfaces. Since the Lexis diagram could have been named for others (Keiding 2011, Vandeschrick 2001), and since we compare with other temporal configurations, let us refer to it as the APC diagram, as seen in Figures **??** and **??**.

The APC diagram in Figure 1 represents years lived on the $y$ axis, calendar years on the $x$ axis, and birth cohorts as the right-ascending diagonals. This is the most common of several possible configurations of the APC dimensions. Individual lifelines (black) are aligned in the birth cohort direction, starting with birth (filled circle) at chronological

Figure 1: An APC diagram with six lifelines.



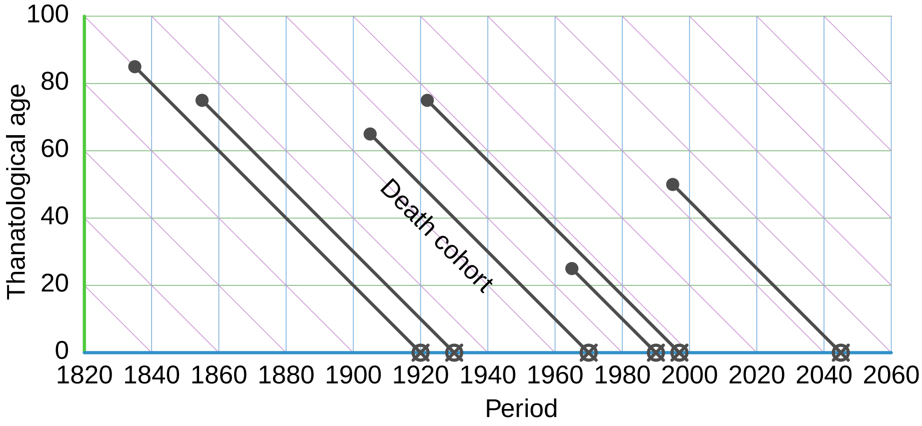

age zero, and death (circled x). Any APC surface can be interpreted along each of these three dimensions of temporal structure.

### 3.3.2 TPD: thanatological age, period, and death cohort

The TPD diagram is best imagined as the inverse of the APC diagram. One may take the same individuals represented in Figure 1 and group them by death cohorts (D) instead of birth cohorts (C). Lifelines then descend such that all endpoints align to thanatological age 0, creating the diagram in Figure 2 in which individuals dying at different ages but in the same time period are grouped together. To our knowledge, the TPD diagram has only appeared once in the literature, as a didactic aid in a proof of symmetry between chronological and thanatological age structure in discrete stationary populations (Villavicencio & Riffe 2016). TPD diagrams may also be useful to arrange events or durations that are logically aligned (or may only be aligned) by time of termination. It may be reasonable to align on termination in cases where this brings preceding patterns of variation into focus.

There are several examples of analysis of this kind of data, usually stemming from a lack of information on chronological age. This is the case, for instance, in biodemographic studies in which wild animals with unknown ages are captured and then followed-up until death Müller et al. (2004; 2007). Other examples are human historical databases, which usually lack information about births, but individuals can be traced from a particular event until death. This is the case in the Barcelona Historical Marriage Database, which collects information about marriage licenses of Barcelona (Spain) from the mid-fifteenth century until the early twentieth century. In this database, ages are unknown, but individuals are first identified in their marriage record and an estimation of the times of death is plausible (Villavicencio et al. 2015). We speculate that TPD diagrams could also be used in biomedical studies for the representation of lifelines preceding deaths from infectious or acquired conditions, when the time of infection or acquisition remains unknown, an issue which has received attention in the statistical literature (see e.g. Chan & Wang

Figure 2: A TPD diagram with six lifelines.



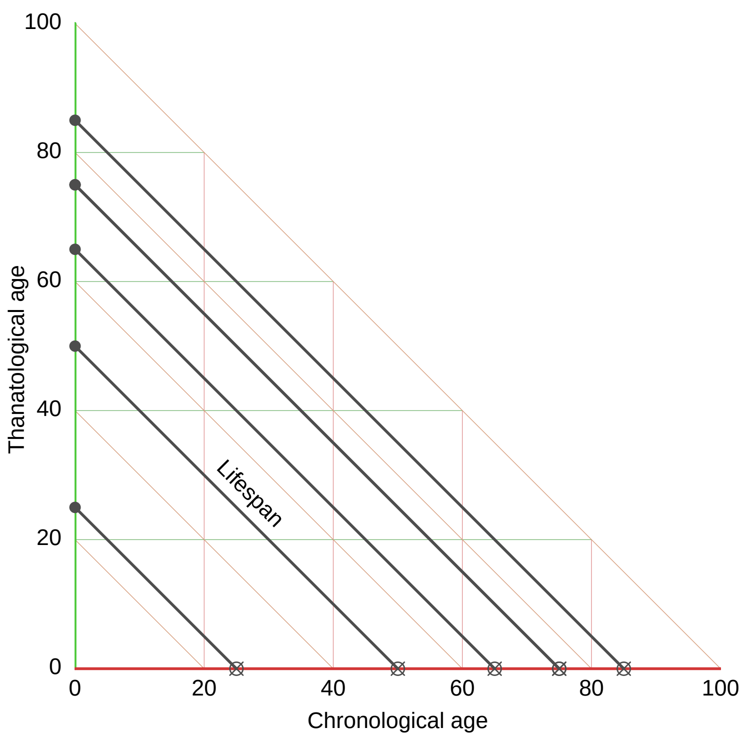

2010).

### 3.3.3 TAL: Thanatological age, chronological age, and lifespan

TAL is an appropriate diagram to examine processes that vary over the lifecourse. More precisely, the TAL plane can highlight variation that is related to time since birth, time until death, length of life, and their combinations. These key aspects of demographic time are compressed to chronological age only in the APC perspective, which can blend out meaningful variation. Since the lifecourse belongs to the cohort perspective, it is best to think of the TAL plane as belonging to some particular birth cohort. Alternatively, a TAL triangle may be taken as a cross-section through the period dimension, a sort of synthetic TAL plane.

To our knowledge, the TAL diagram has only appeared once in the literature, in an exploration and classification of late-life health conditions (Riffe et al. 2015). There are however instances of statistical designs adapted to this coordinate plane (see e.g., Jewell 2016, Dempsey & McCullagh 2016). The TAL diagram in Figure 3 contains no indication of period or cohorts, as calendar time is blended out in this diagram. The lifelines depicted are identical to those shown in APC Figure 1 and TPD Figure 2. The TAL diagram is useful for characterizing patterns of prevalence of health conditions. We speculate that data structured and aligned in this way may yield hitherto undescribed patterns in other contexts, such as pregnancy (from the perspective of either mother or fetus) by time since conception, time until parturition, and length of gestation, or even smaller event-history time scales, or patterns of growth or reproduction in non-human species.

Figure 3: A TAL diagram with six lifelines (*).

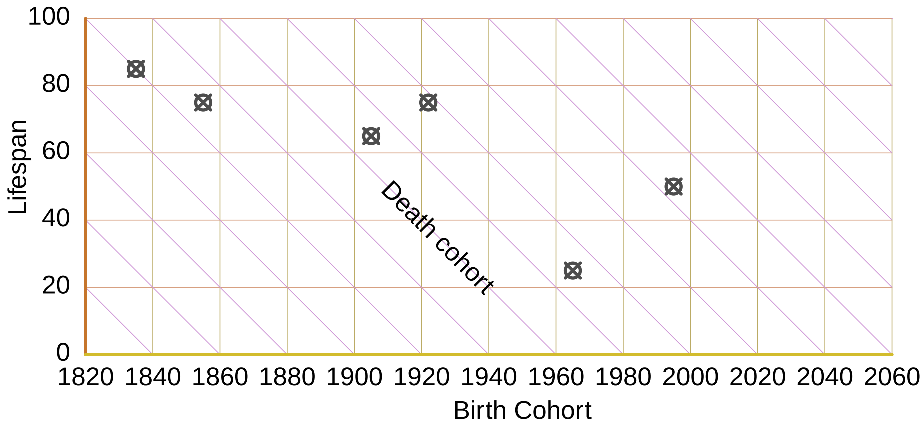



(*) Since two of the six lifelines are of equal length (75), they are overlapped in this figure and appear to be five.

### 3.3.4 LCD: Lifespan, birth cohort, and death cohort

The LCD diagram completes our set of identities. It is based on the relationship between lifespan, birth cohort, and death cohort. In Figure 4, lifespans are indexed by the $y$-axis, while birth cohorts are indexed by the $x$-axis, and death cohorts are found in descending diagonals. To structure data on these three time measures implies excluding time-varying information over the lifecourse. An individual only ever has one lifespan, one birth cohort, and one death cohort, such that the LCD coordinates of an individual are constant throughout life. The LCD plane is therefore orthogonal to lifelines, and individuals are located with points, rather than life segments. In Figure 4, the same six individuals from previous diagram figures are represented with crossed circles.

Figure 4: An LCD diagram in two projections.




We recommend this mapping for plotting surfaces of values that are cumulative or static over the lifecourse, but that may vary over time or by length of life. Imagine an LCD surface of cumulative lifecourse consumptive surplus or deficit, or anything else that might vary by lifespan and moment of birth or death, such as children ever born, years of retirement, the size of trees or other aspects of forestry, populations of buildings in large cities, and so forth. Lexis (1875) describes an analogous relationship between marriage cohort, separation cohort, and duration of marriage.

# 4 A general relationship between events and durations

The four identity-based diagrams discussed in prior sections are likely straightforward, either because the Lexis diagram is already familair to the reader, or because Cartesian representations are widely used. However, the special relationship between these diagrams is based on a single hexad identity, which is less straightforward, and its resultant diagram is best derived from a more general groundwork. In this section we therefore describe a more general approach to understanding and constructing higher order temporal identities. This approach is based on a categorization of time measures into events and durations, and the realization that durations derive from events in calendar time.

The general relationship between events and durations serves not only to introduce the full demgographic time framework, but also to compare it with other relatively complicated temporal designs in the literature. Each of the six time measures that we have treated can be categorized into two basic types: Events and durations. Events include birth (C) and death (D), as well as period itself (P). Durations are time differences between pairs of events. Chronological age ($A = P - C$), thanatological age ($T = D - P$), and lifespan ($L = D - C$) are in this senese durations.

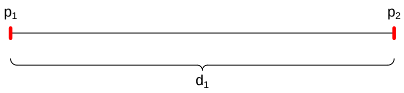
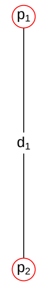
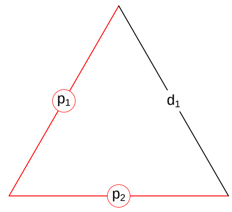
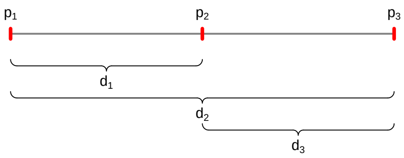
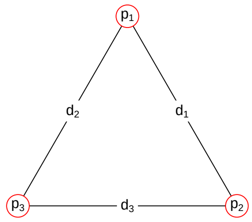
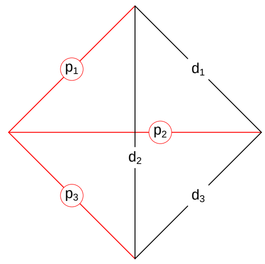
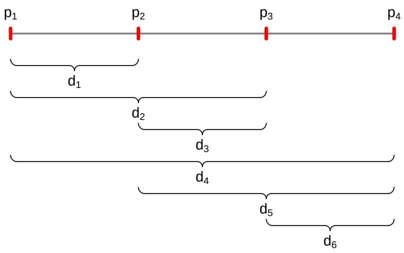
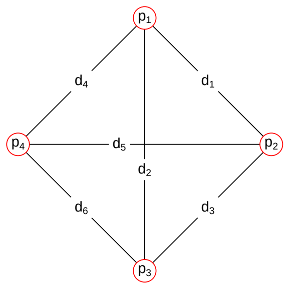
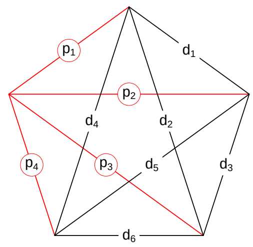
Let us formalize the notions of $n$ points in time $\mathbf{p} = [p_1 \dots p_n]$ and the $m$ durations $\mathbf{d} = [d_1 \dots d_m]$ between these points. $\mathbf{d}$ is linearly dependant on $\mathbf{p}$ in the sense that a linear transformation of $\mathbf{p}$ yields $\mathbf{d}$. A set of $n$ events implies a total of $m = \binom{n}{2} = n(n-1)/2$ durations. The total number of time measures implied by a set of $n$ events is therefore $n + m$. A set of time measures derived in this way, $\mathbf{g} = [\mathbf{p}, \mathbf{d}]$, forms an identity. In the same way that $\mathbf{d}$ derives from $\mathbf{p}$, there are $(n+1)^{(n-1)}$ ways to complete $\mathbf{g}$ from a size-$n$ set of potentially mixed durations and events. Proofs for these statements are found in Appendix A.

The relationship between events and durations can be systematically represented in a series of timelines and graphs that may better guide intuition. Table 3 displays a timeline and two levels of graph representation for two, three, and four event sets. The left column shows timelines, a familiar linear representation of time, with events marked with red ticks labelled with $p_1 \dots p_n$. Durations span each of the $m$ possible event dyads and are drawn below the main timeline as curly braces labelled with $d_1 \dots d_m$. It is evident, that as $n$ increases, $m$ increases relatively faster, and this timeline graphical representation is not efficient for $n \gtrsim 4$. Drawing events on a timeline implies an ordering, but the events of $\mathbf{p}$ need not be ordered in any particular way in order for these relationships to hold.

The joint relationship between durations and events is more explicit and more compact in a graph representation. The middle column of Table 3, labelled *graph* consists in the simplest graph representation of the event-duration timeline, with $n$ red vertices for each of the events in $\mathbf{p}$. A full connection of the graph yields $m$ edges for each of the possible durations in $\mathbf{d}$. This graph representation loses the linear time analogy of the timeline, but makes the depdenency of $\mathbf{d}$ on $\mathbf{p}$, and other interdependencies between time measures more explicit.

The right column of Table 3, labelled *temporal plane graph* redraws the graph with a total of $n + 1$ vertices and $m + n$ edges for the elements of both $\mathbf{d}$ and $\mathbf{p}$. All events of $\mathbf{p}$ connect to a single vertex. Event edges are indicated in red with red-circled labels and nodes have no direct time-measure meaning. In this rendering, each triangle formed by three mutually connecting edges represents a triad identity. The top row $n = 2$ consists in a single identity. Three and four events imply a total of four and ten triad identies, respectively, and in general a given higher order identity will yield $\binom{n+1}{3}$ triad identities. We call this a temporal plane graph because the triangle resulting from any given triad sub-identity can be extended over all valid values of its time measures to form a temporal plane, as of the diagrams in Section 3.3.

Table 3: Event-duration timeline, graph, and temporal plane graphs for two, three, and four event sequences.

| nr. events | timeline | graph | temporal plane graph |
|---|---|---|---|
| $n = 2$ |  |  |  |
| $n = 3$ |  |  |  |
| $n = 4$ |  |  |  |

## 4.1 Examples

Some brief examples will add intuition to the interpretation of Table 3.

**Example 1: The Lexis surface** Let $\mathbf{p}$ have two elements, as in the first row of Table 3. Then $\mathbf{d}$ consists of just one element, defined as

$$d_1 = p_2 - p_1 \qquad . \tag{1}$$

Interpreting $d_1$ as *age*, $p_2$ as *period*, and $p_1$ as *birth cohort* yields the APC identity. The Lexis surface is defined as the plane of all possible combinations between *age* and *period*. with cohorts as diagonals results from the transformation $P^2 \rightarrow M^2$ as shown to be possible in theorem 3.

**Example 2: Lexis' marriage identity** Along with his well known 2-dimensional diagram Lexis (1875) also described a 3-dimensional extension applied to the marriage and separation processes.

Let $\mathbf{p}$ have three elements, as in the second row of Table 3. Then $\mathbf{d}$ is defined as

$$\begin{aligned} d_1 &= p_2 - p_1 \\ d_2 &= p_3 - p_1 \\ d_3 &= p_3 - p_2 \end{aligned} \quad . \tag{2}$$

Interpreting $p_1$ as *birth cohort*, $p_2$ as *marriage cohort* and $p_3$ as *separation cohort* yields the durations $d_1$ as *age at marriage*, $d_2$ as *age at separation*, and $d_3$ as *duration of marriage*. Lexis' "marriage space" is reconstructed by transforming $P^3 \rightarrow M^3$ where $M$ has three orthogonal basis vectors corresponding to $(p_1, d_1, d_3)$. also need to work out that theorem for this. also, is period implied by the ages? if so, is this actually a 4-event identity?

**Example 3: Adding death cohort to the Lexis surface** As in Example 2 we start with a three element vector $\mathbf{p}$ yielding the very same identities as in equation set (3) and the second row of Table 3, but with different interpretations. Interpreting $p_1$ as *birth cohort*, $p_2$ as *period* and $p_3$ as *death cohort* yields the durations $d_1$ as *chronological age*, $d_2$ as *lifespan*, and $d_3$ as *time to death*. This vector space $P^3$ contains the Lexis surface as a sub-space, as well as the other planes presented in Section 3.3. We return to this identity in the following sections.

**Example 4: Brinks' Illness-Death model** Brinks et al. (2014) describes an illness-death process atop the Lexis surface, and with diagnosis and death as additional events, for a total of four events. Let $\mathbf{p}$ have four elements, as in the third row of Table 3. Then $\mathbf{d}$ is defined as:

$$\begin{aligned} d_1 &= p_2 - p_1 \\ d_2 &= p_3 - p_1 \\ d_3 &= p_3 - p_2 \\ d_4 &= p_4 - p_1 \\ d_5 &= p_4 - p_2 \\ d_6 &= p_4 - p_3 \end{aligned} \quad . \tag{3}$$

Interpreting $p_1$ as *birth cohort*, $p_2$ as *period*, $p_3$ as diagnosis, and $p_4$ as *death cohort* yields the following composition of $\mathbf{d}$: $d_1$ is *chronological age*, $d_2$ is *age at diagnosis*, $d_3$ is *time to diagnosis*,[3] $d_4$ is *lifespan*, $d_5$ is *time to death*, and $d_6$ is duration of illness (an irreversible state). The complete identity of this model therefore implies a total of ten time measures, although not all measures are explicitly mentioned or used in the model of Brinks et al. (2014).

# 5 A tetrahedron relates the six time indices.

The demographic time framework we present includes three events (period, birth cohort, and death cohort), and it therefore leads to a temporal plane graph based on the second row and third column of Table 3, here redrawn in Figure 5a withg edges labelled by the six time measures and colored consistent previous figures. A slight rearrangement of

[3] Time to diagnosis can also be time *since* diagnosis if the elements of $\mathbf{p}$ are rearranged. Note that the elements of $\mathbf{p}$ need not be ordered. Although birth and death are likely the first and last elements of $\mathbf{p}$, this is not strictly required.

the vertices in Figure 5b yields a graph representation of the tetrahedron, the simplest platonic solid.

An examination of either layout in Figure 5

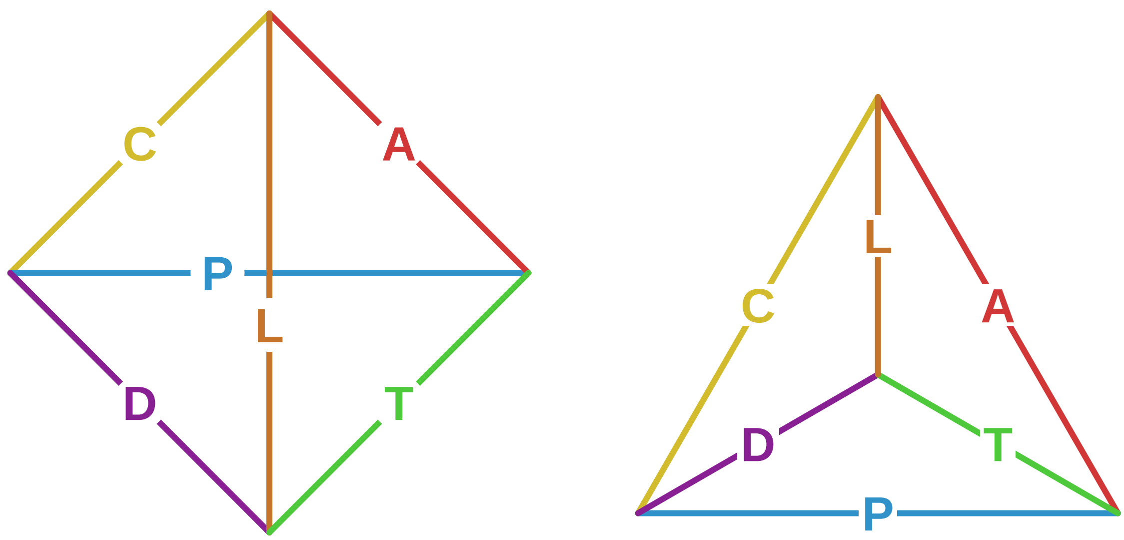
Each of the four triad identities may be thought of as a two-dimensional plane fully defined by any two of its three constituent time indices. In this case, we may imagine any of the excluded time measures as capable of providing depth, a potential $z$-coordinate, for the sake of a mental image. Having a non-redundant third dimension implies a multitude of parallel planes for the given triad identity, each plane belonging to a unique value of the third time dimension. Any of the identities can be extended in this way to fill a space. A space derived by extending any of the triad identities into its lacking dimension implies each of the other triad identities, making a total of six time indices. In essence, the four triad identities may be thought of as the four faces of a tetrahedron. If an additional time measure is added to any face (triad identity), the six demographic time indices can be derived, matching the six edges of the tetrahedron. This three-dimensional construct unifies the six indices of demographic time, and is the subject of this paper.

Let us first more rigorously define the previously-mentioned tetrahedron. Luckily, the edges and vertices of a tetrahedron are easily rendered in a two-dimensional graph, as seen in Figure 5, with vertices labeled in black and the six time indices colored following the pattern from Table 2. The tetrahedron is composed with the APC plane at the base and vertex 4 on top. The same graph could be composed in four basic ways, depending on which identity forms the base.

Figure 5: Graphs of demographic time hexad identity, with edges labeled by the six time indices.

(a) vertices on the unit circle

(b) vertices in tetrahedral layout




The edges APC at the base define the much-studied APC plane. If the only information we have is chronological age, period, and birth cohort (or just two of these), then we have no access to the vertex 4. Each of the faces of the tetrahedron has this quality. The South face TPD has no access to 1. The Northeast face, TAL has no connection to 3, and the Northwest face LCD lacks a connection to 2. The four triad identities that make up the faces of the tetrahedron are stuck in “flatland” and do not yield the full 3d space.

However, the 16 other possible combinations of three time indices will recreate the full tetrahedron (hexad identity).

# 6 Diagram of the hexad identity

There are different ways to proportion this three dimensional construct, of which we only present the isotropric mapping. In an isotropic projection, the tetrahedron is regular, such that all edges are of the same length, and the units of each of the six represented time measures are therefore equal. In this case, the four triad identities map to temporal planes as tessellations of equilateral triangles, and the four planes are joined together such that each is parallel to a face from the regular tetrahedron. When the plane parallel to each respective face is repeated in equal intervals, we have an isotropic 3d space.[4] Displaying all planes simultaneously creates a very dense and difficult-to-read diagram. We opt to delineate the space using particular planes and intersections.

Figure 6 gives a view of a demographic time diagram that corresponds to the hexad identity, where birth-cohort TAL cross-sectional planes are placed in sequence in a perspective drawing.[5] The most recent TAL plane, for the year 2000, is placed in the front, whereas past TAL planes are stacked behind it, highlighted in 25-year intervals. The left edge of the frontmost TAL plane is labelled as an axis for thanatological age, although the same tick marks also serve for completed lifespan. The base of this figure is the APC plane, drawn for thanatological age 0. Each of the TAL planes sits atop a single birth cohort line from the familiar APC plane that makes up the base of the figure.

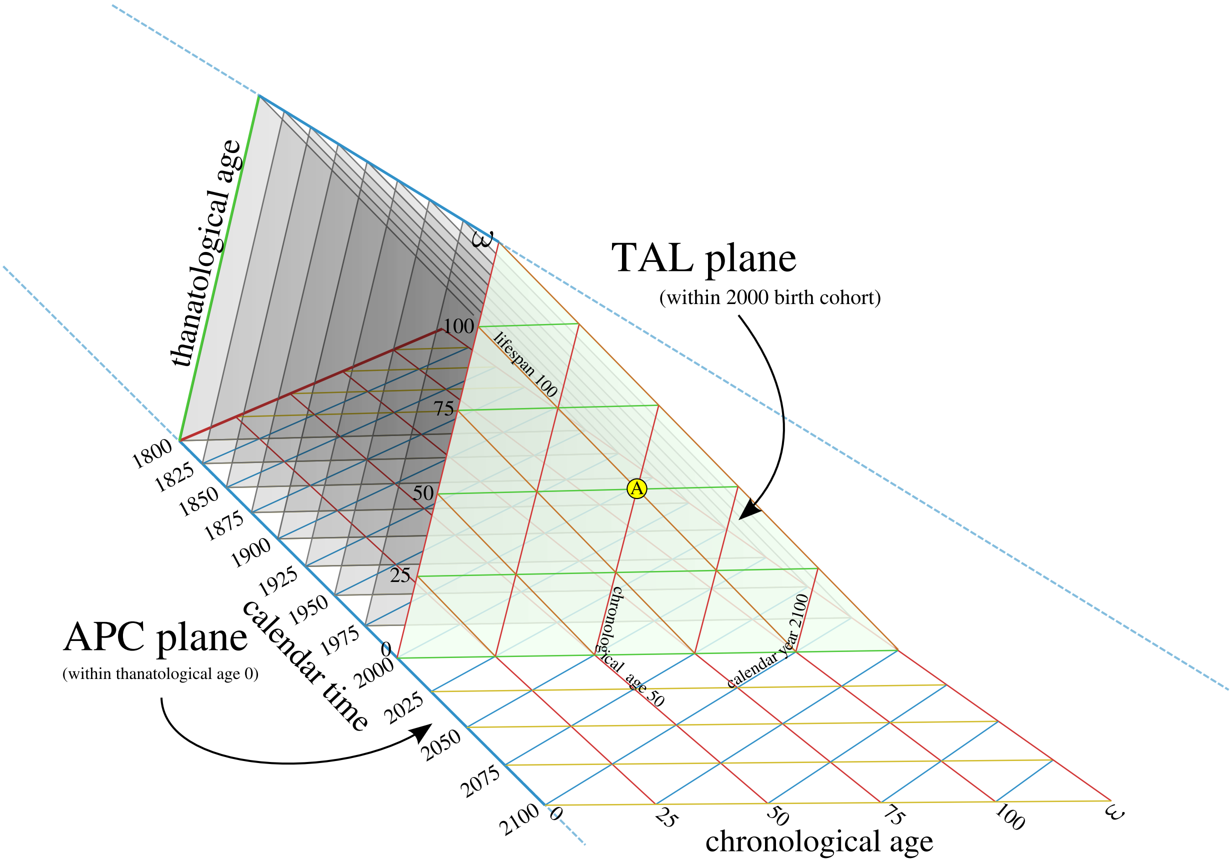
For example, imagine an infant born in the year 2000. Without further information, we only know that this infant is located somewhere on the thanatological age axis of the front TAL plane. If this infant is destined to die in the year 2100, then the vertical position at birth will be at the axis tick for thanatological age 100. This person's entire life stays on the 100 lifespan line (labelled), descending over time towards thanatological age 0. Point A marks the midpoint in life for this individual, at chronological age 50 (red line, labelled), and thanatological age 50 (green line). If another APC plane were drawn through thanatological age 50, we would see that point A is in the year 2050. Since all individuals born in the year 2000 complete the same age in the same year, we can also recuperate the year for point A by following the chronological age 50 line (red) down to where it meets the blue line for the year 2050. The lifeline descends downward toward the APC plane for thanatological age 0 at chronological age 100, meeting the year 2100, which determines the death cohort.

The density and location of imaginary lifelines in this diagram, omitting migration, is purely a function of birth cohort size and survival. For extinct cohorts all lifelines can be positioned, but for the 2000 birth cohort this is not yet the case. Most of the front TAL plane is in the future. One may imagine yet another plane intersecting this space— the "present plane", which is identical to the period TAL plane for the present

[4]The isotropic space that results from this framework is known in other disciplines with different nomenclatures. In geometry, this structure is called the tetrahedral-octahedral honeycomb, a variety of space-filling tessellation. In architecture, it is found in the octet truss system. In physics it is called the isotropic vector matrix. Constructs following this geometry exist in nature, in other theoretical settings, and in man-made structures.

[5]The coordinates used to render Figure 6 are isotropic. However, there are no 60° angles in this figure due to the use of parallax and an indirect viewing angle in this rendering for the sake of increased legibility.

Figure 6: Diagram of the hexad identity, showing a sequence of TAL planes intersecting with a single APC plane.




moment. To see how this plane divides the space, imagine that we are in the year 2025, and follow the blue line in the APC base inward 25 years to where it meets the red line for chronological age 25, and follow the red line up the front TAL plane. A single plane cuts through the year 2025 and chronological age 25 from the year 2000 birth cohort. This plane shifts forward or backward in time to meet the present year. In this particular plane, the coordinates T, L, and D are uncertain. The period TAL plane $\omega$ years in the past is fully identified, ergo, theoretically the lifespan of each individual in the time of Lexis is knowable.

Figure 6 could have been drawn with TPD or LCD planes highlighted as well, but these can still be imagined upon the current rendering. TPD planes transect this space through any given chronological age, for instance. Imagine a wall on the left side of the prism, cutting through chronological age 0 (recall Figure 2). In this case, the thanatological age axis is indicated in the very back of the diagram, calendar time becomes another axis, and death cohort diagonals are not drawn. TPD planes sequence inward from this first plane, always forming cross-sections through chronological age. The LCD plane is to be found by rotating the current prism such that the angle of view is directly down (or up) lifelines, which would then appear as points (recall Figure 4).

Further, the two classes of planes drawn (APC and TAL) could have been drawn and labelled differently. In practice, the diagram of any triad identity may be drawn held constant for any one of its three missing indices with no loss of generality. For example, we have mentioned period (P) and birth cohort (C) TAL planes, but there must also be a third TAL plane that is held constant for death cohort (D). Another example of such multiplicity is found in the birth cohort TAL diagram. In this diagram, lifespan lines can also be interpreted as death cohorts (D), chronological age lines are also period lines (P), and thanatological age (T) is maintained. This means that the birth cohort TAL diagram is also a birth cohort TPD diagram. Multiplicity of this kind can be systematically confirmed, since the C time measure is not part of TPD. To not overly extend this exposition, and to avoid undue confusion, we neither delineate every possible cross-section nor the possible interpretations of each cross-section. However, in general there are four total ways to cut the space that are parallel to the faces of the hexad identity, and each cut has three possible diagram interpretations.

The essential property of this perspective diagram is that lifelines start and end in parallel, desceding downward and forward in time. A real population of renewing lives, spread over time and over the typical range of human lifespans, will tend to fill the entirety of the prism depicted in Figure 6, and any given point in the prism can be given six demographic time coordinates, of which two or three are redundant.

# 7 Application

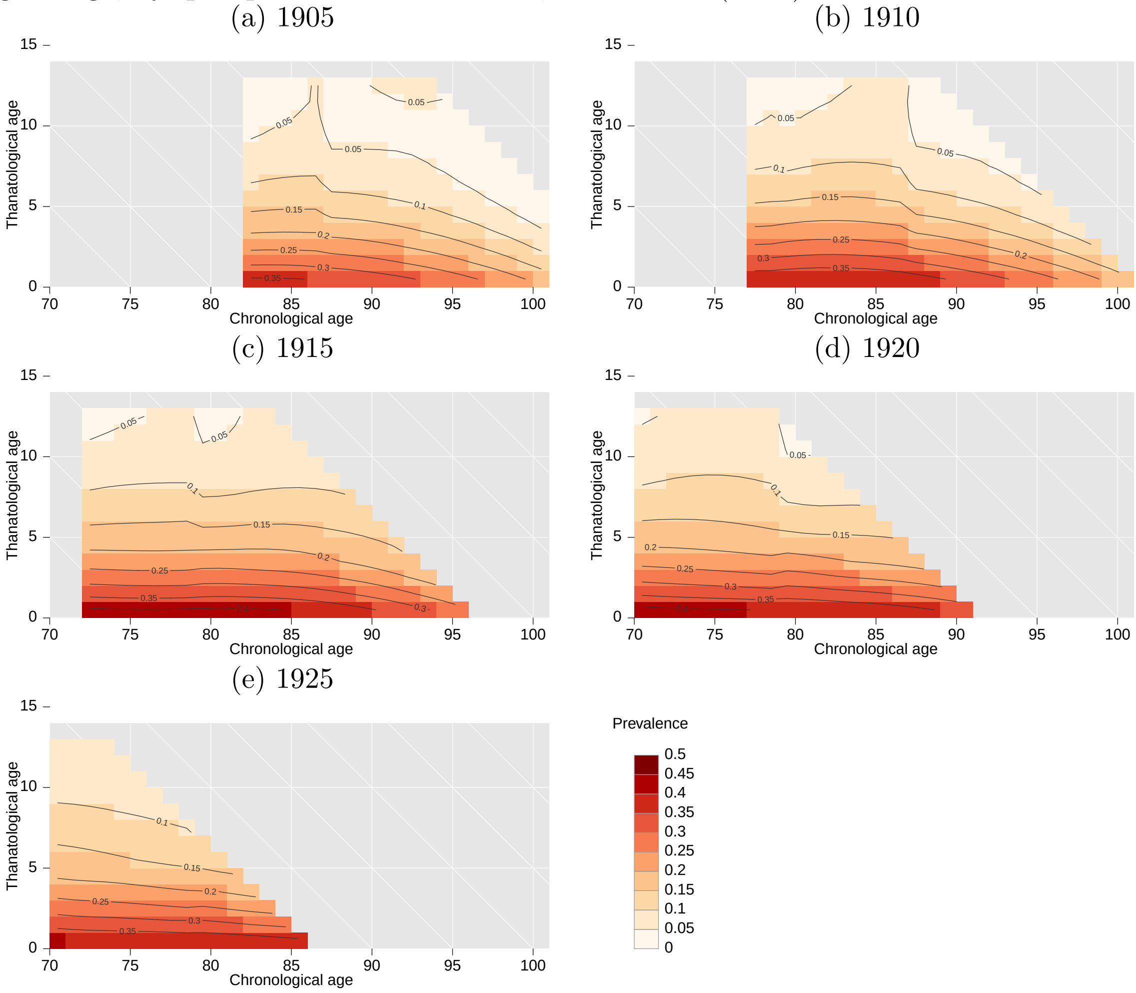
The relationship between the six measures of demographic time is true in the same sense that mathematics is true: Under linear time it is an internally valid set of relationships, and this is self-evident. We have mentioned that the coordinates described here may be useful for the visualization of data, to enable discovery, and to better inform demographic methods. We have not yet mentioned how such developments might arise in practice. We therefore give a schematic overview of our own process of scientific inquiry and reflection that was based this coordinate system, and that would not have arisen without it. This chain of inquiry is meant to demonstrate the usefulness of the present framework, but it is far from an exhaustive application of its potential for other substantive questions, nor is the case study described in complete rigor. Specifically, we reason that projections or comparisons of healthy life expectancy (HLE) are in many cases biased in period prevalence-based models unless one takes into account the thanatological age pattern of prevalence, as well as mortality differences.

There are three steps in our empirical inquiry. The first step is to visualize variables on health outcomes using our framework. The second step is to assess the primary time measures over which health outcomes appear to vary. Under the assumption that these patterns of temporal variation are empirically regular, we proceed to develop a method of standardizing health expectancy calculations for morbidity conditions whose prevalence is more closely related to thanatological age. Finally, one can reason that period estimates of health expectancies for certain health conditions are biased when mortality has been or will-be changing, and comparisons of HLE between populations with different mortality are also biased. We conclude that comparisons of health expectancies might be biased in ways not previously documented.

Let us take the example of self-reported health (SRH). The data come from the RAND version of the US Health and Retirement Study (RAND 2013). Since this survey has multiple observations of individuals, as well as a mortality follow-up, we have each of the six time measures for each observation. Further methodological details are given by Riffe et al. (2015).

Figure 7 displays a series of TAL surface plots of SRH prevalance, each referring to a different quinquennial birth cohort (1905-1909, etc). The TAL surfaces for each successive

Figure 7: Prevalence of males self-reporting poor health by chronological and thanatological age, by quinquennial birth cohorts, 1905-1925. (HRS)




birth cohort are shifted by five years because the observation window available is from 1992 to 2011 for each cohort. Contour lines in the surfaces indicate the primary direction of variation. Downward diagonals indicate lifespans, which the reader may also think of as very specific birth-death cohorts. These are the diagonals along which lifelines may be imagined, as suggested in Figures 3 and 6. For each of these birth-death cohorts we have a prevalence trajectory— empirical examples of the lifeline morbidity trajectories often conceptually diagrammed in the literature on morbidity compression (e.g., Fries 2005). In each surface the primary direction of variation is along thanatological age, and not chronological age. The prevalence for those with $t$ remaining years of life is similar in these data, irrespective of chronological age, birth cohort, or ultimate lifespan.

The marginal chronological age pattern of SRH, as measured here (the Sullivan curve, (Sullivan 1971)), shows an increasing tendency over age. However, such an increasing line is a marginal ruse, due to an interaction between the distribution of lifespans and the relatively fixed underlying pattern of morbidity seen in Figure 7. These surfaces can indeed be tidily summarized with a single line, but it is a line over the thanatological age margin rather than over chronological age.

Since the patterns for each of these cohorts can be presumed to be the same, any shifting in the distribution of lifespan ought not produce a change in the expected years of poor health for a given lifespan. Further, the life years spent in poor health should also be approximately the same "on average", even if the underlying mortality patterns shift. If morbidity change is a pure function of thanatological age, an increase in life expectancy should increase healthy life expectancy by the same amount. This is not the prediction when we base analyses on the chronological age pattern of self-reported health. An underlying morbidity pattern this stable would predict improvements in the marginal chronological age pattern of self-reported health if the lifespan distribution were to shift to higher ages. This bias in the current status quo of morbidity measurement and prediction leads to pessimistic morbidity scenarios when mortality improvements are foreseen, and it undermines health expectancy comparisons between groups with different mortality (Van Raalte & Riffe 2016). Cohort health expectancies are in either case unbiased, but these are also not common.

Using the data from our example surfaces, we can calculate an average prevalence trajectory with the approach to death and calculate some basic results that support our case. Let us take the population of US males aged 60 and older, and assume that the trajectory derived from the Figure 7 surfaces is valid for them. If we apply this trajectory to the synthetic stationary population of each year from 1980 and 2010 (HMD 2016), we can calculate the resulting healthy and unhealthy life expectancies, and compare these with the expectancies that we would have projected assuming the 1980 Sullivan curve. Total remaining life expectancy at age 60 increased 4.3 years from 17.4 to 21.7 years from 1980 to 2010. Assuming the thanatological trajectory of morbidity, we calculate healthy life expectancies of 15.7 and 19.9, respectively, an increase of 4.2 years. Unhealthy life expectancy in this scenario increased just 0.1 years. Had we used the Sullivan curve from 1980 to calculate the 2010 values, we would have predicted an increase of 0.7 years in unhealthy life expectancy, or 39% versus the 4% "observed" in this exercise.

This is a large difference in projected morbidity, and it is based on a relatively minor tweak to standard methodology, itself inspired by viewing data under the conditions enabled by this temporal framework and adjusting standard demographic methods to capture the direction of temporal variation in data. There is a wide variety of prevalence patterns when viewed in this way (Riffe et al. 2015, Wolf et al. 2015), and much empirical and methodological work is still required to verify that these findings are representative and to understand the consequences for the standard ways of comparing and projecting HLE.

Our objective in this application has been to demonstrate how viewing data under the rigorous conditions enabled by the time-framework we propose can lead to new scientific understandings of processes over the life course. Applications for other aspects of the lifecourse, particular stages of the lifecourse, non-human species on all time scales, populations of inanimate items on all time scales, and myriad other substantive areas may gain new insight by applying the relationships contained in the demographic time identity or its consituent triad identities.

# 8 Discussion

In this paper we describe a relationship between six different measures of demographic time: chronological age, period, birth cohort, time to death, death cohort, and individual lifespan. In Section 3 we show how combinations of these time measures imply four triad identities. Each triad identity consists in simple linear relationship between its three constituent time measures. We describe how each triad identity can be extended into a temporal plane, with a characteristic diagram. The four triad identities underly a family of four diagrams. These diagrams include the familiar Lexis diagram (Sec 3.3.1), but also three either new or uncommon diagrams. The TPD diagram (Sec. 3.3.2) relates time of death, period, and time until death, and is therefore a sort of dual to the Lexis diagram. The TAL diagram (Sec. 3.3.3) relates the three duration measures of time to death, chronological age, and lifespan, and it may be useful for analyzing patterns that vary over the lifecourse and/or by length of life. The CDL diagram (Sec. **??**) relates time of birth with time of death and lifespan, and it may be useful for analyzing patterns that vary over time and by length of life, but not necessarily over the lifecourse. These four identities and diagrams relate to one another in a single relationship.

To present this single relationship with some rigor, in Section 4 we digress to present a more general event-duration framework. These general terms allow us to present the demographic time hexad identity as a special case of the event-duration framework in Section 4.1. We compare this identity with other relatively complicated temporal relationships in the literature, including the Lexis (1875) marriage identity and the illness-death model of **?**. In Section 5 we describe how the graph of the demographic time identity is also the graph of a tetrahedron. As a three-dimensional solid, the tetrahedron forms the basis of the three dimensional extension of the demograhic time identity. Each of the four faces of the tetrahedron is parallel to one of the four temporal planes.

In Section 6 we render a diagram of the demographic time identity. We argue that the full three dimensional diagram is not necesarily a practical way to represent demographic data, but that it forms a useful reference to understand demographic processes. In general, data structured by all six demographic time measures can be represented on any of the four diagrams if controlled properly. In Section 7 we present a brief application of this technique to the prevalence of poor self-reported health in older ages in the United States. Prevalance data are displayed in a series of TAL surfaces (See Sec. 3.3.3) representing the end-of-life experience of a sequence of five birth cohorts (1905-1925). Together, these surfaces represent a partial filling of the three-dimensional diagram drawn in Figure 6. As such, this prevalence data is structured by all six demographic time measures.

The contemporary practice of (macro) demography is based on the premise that vital rates, and other kinds of rates over the lifecourse, are the truest measure of demographic forces. Rates are paramount because they tend to vary in empirically regular ways over the life course. The scalings and movements of primary vital rates fall within a limited range for humans. For this reason, many of the methods of demography are developed to estimate rates, independent of population composition, or to partition crude magnitudes into the effects of population age structure and pure vital rates. Controlling for age like this is in a more general sense controlling for temporal variation in stocks. To the extent that regular temporal variation relates to the end of life, or the length of life, common age-standardization does not fully account for such structure.

The techniques used to age-standardize mortality and fertility estimation are at times applied to other kinds of transitions over the life course. For example, one may estimate an age pattern to some degenerative disease, or the ability to carry out some common activities of daily life. However, much of the regular temporal variation for the prevalence of such conditions is by time-to-death or lifespan, rather than by chronological age. Apparent chronological age patterns in such cases are artifactual and do not represent the same kind of intrinsic meaning as does the *age pattern* of mortality. Further kinds of temporal standardization must be developed in order to measure and understand the natural patterns of such conditions over the lifecourse. The measurement of such conditions may benefit from consideration of the framework presented in this paper. To this end, Table **??** provides all combinations of information that are sufficient to derive the full set of six time measures. Panel surveys with mortality follow-ups already provide the requisite information, as do linkable registers that include items such as health measures or proxies and relevant dates of birth, observation, and death. Other kinds of populations, such as animals or items, may have quite different data-gathering mechanisms.

An effective way to detect patterns in temporal variation is via data visualization. The coordinate system proposed in this paper is conceived as one adequate to capture such variation, and we suggest its use for visualizing data, probably via small multiples of successive time slices parallel to any of the four triad identities, similar to that shown in Figure 7. Such visualization strategies at this time are exploratory, and this is a technique that may benefit from further refinement. Further, a cross-section through the demographic time-space need not be parallel to one of the four identity-planes.

Mortality determines three of the dimensions of demographic time, and it therefore makes little sense to model mortality using all six time measures. Rather, mortality determines the placement of lifelines in the three-dimensional space. Death itself is only representable with the endpoints of lifelines, ergo the APC plane through thanatological age zero. This full coordinate system instead defines and contains the space through which life is lived; any of the six measures may be pertinent in the case of conditions and states that vary over and within the lifecourse. An obvious application for the present model, given data commonly (and publicly) available at this time, are late-life health conditions, although there may be other substantive areas of application.

Furthermore, we believe in the pedagogical value of the framework introduced in this paper. We hope that the present inquiry will be useful as a teaching instrument in the same way as APC diagrams have formed a part of basic demographic education. The relationship between the six dimensions of demographic time helps situate the APC paradigm in a broader framework. Just as scientific discovery in general depends partly on the development of finer optics and instrumentation, we hope that the framework we describe will prove an instrument to enable new discoveries in formal, and empirical demography, as well as other diverse fields of investigation.

# References

Aigner, M., Ziegler, G. M., & Quarteroni, A. (2010). *Proofs from the book* (Vol. 274). Springer.

Brinks, R., Landwehr, S., Fischer-Betz, R., Schneider, M., & Giani, G. (2014). Lexis diagram and illness-death model: Simulating populations in chronic disease epidemiology. *PLoS ONE*, *9*(9), e106043.

Brouard, N. (1986). Structure et dynamique des populations: La pyramide des années à vivre, aspects nationaux et exemples régionaux. *Espace, populations, sociétés*, *4*(2), 157–168.

Caselli, G., Vallin, J., & Wunsch, G. (Eds.). (2006). *Demography: Analysis and synthesis* (Vol. 1). Academic Press.

Cayley, A. (1889). A theorem on trees. *Quarterly Journal of Pure Applied Mathematics*, *23*, 376-378.

Chan, K. C. G., & Wang, M.-C. (2010). Backward estimation of stochastic processes with failure events as time origins. *The Annals of Applied Statistics*, *4*(3), 1602–1620.

Dempsey, W., & McCullagh, P. (2016). *Survival models and health consequences.* (Unpublished. Available at https://arxiv.org/pdf/1301.2699v3.pdf (retrieved on 4 April 2017))

Fries, J. F. (2005). Frailty, heart disease, and stroke: The compression of morbidity paradigm. *American Journal of Preventive Medicine*, *29*(5), 164–168.

Human Mortality Database. (n.d.). (University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. Available at http://www.mortality.org (data downloaded on 10 July 2014))

Jewell, N. P. (2016). Natural history of diseases: Statistical designs and issues. *Clinical Pharmacology and Therapeutics*, *100*(4), 353–361.

Keiding, N. (2011). Age-period-cohort analysis in the 1870s: Diagrams, stereograms, and the basic differential equation. *Canadian Journal of Statistics*, *39*(3), 405–420.

Knapp, G. F. (1868). *Über die Ermittlung der Sterblichkeit aus den Aufzeichnungen der Bevölkerungs-Statistik*. Leipzig: J. C. Hinrich.

Lexis, W. (1875). *Einleitung in die Theorie der Bevölkerungsstatistik*. Strasbourg: K. J. Trübner. (Available at https://archive.org/details/einleitungindie00lexigoog (retrieved on 27 March 2016))

Müller, H.-G., Wang, J.-L., Carey, J. R., Caswell-Chen, E. P., Chen, C., Papadopoulos, N., & Yao, F. (2004). Demographic window to aging in the wild: constructing life tables and estimating survival functions from marked individuals of unknown age. *Aging Cell*, *3*(3), 125–131.

Müller, H.-G., Wang, J.-L., Yu, W., Delaigle, A., & Carey, J. R. (2007). Survival and aging in the wild via residual demography. *Theoretical Population Biology*, *72*(4), 513–522.

Perozzo, L. (1880). Della rappresentazione graphica di una collettività di individui nella successione del tempo, e in particolare dei diagrammi a tre coordinate. *Annali di Statistica (Serie 2ª)*, *12*, 1–16. (Available at https://archive.org/details/annalidistatist21commgoog (retrieved on 22 July 2016))

Pressat, R. (1961). L'analyse démographique: méthodes, résultats, applications. *Population*, *16*(3), 505–508.

RAND. (2013). *RAND HRS Data, Version M.* Santa Monica, CA. (Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration.)

Riffe, T., Chung, P. H., Spijker, J., & MacInnes, J. (2015). *Time-to-death patterns in markers of age and dependency.* (MPIDR Working Paper WP-2015-003. Rostock: Max Planck Institute for Demographic Research)

Sullivan, D. (1971). A single index of mortality and morbidity. *HSMHA Health Reports*, *86*(4), 347-354.

Vandeschrick, C. (2001). The lexis diagram, a misnomer. *Demographic Research*, *4*(3), 97–124.

Van Raalte, A., & Riffe, T. (2016). *Accounting for temporal variation in morbidity.* (Paper presented at Population Association of America 2016 Annual Meeting, Washington, D. C., USA, Mar 31–Apr 2, 2016)

Villavicencio, F., Jordà, J. P., & Pujadas-Mora, J. M. (2015). Reconstructing lifespans through historical marriage records of Barcelona from the sixteenth and seventeenth centuries. In G. Bloothooft, P. Christen, K. Mandemakers, & M. Schraagen (Eds.), *Population reconstruction* (pp. 199–216). Cham: Springer.

Villavicencio, F., & Riffe, T. (2016). Symmetries between life lived and left in finite stationary populations. *Demographic Research*, *35*(14), 381–398.

Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D. A., Boe, C., Bubenheim, M., ... Vachon, P. (2007). *Methods protocol for the Human Mortality Database* (Tech. Rep.). University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. (Available at http://www.mortality.org/Public/Docs/MethodsProtocol.pdf (retrieved on 11 June 2015))

Wolf, D. A., Freedman, V. A., Ondrich, J. I., Seplaki, C. L., & Spillman, B. C. (2015). Disability trajectories at the end of life: A "countdown" model. *The Journals of Gerontology: Series B (Psychological Sciences and Social Sciences)*, *70*(5), 745–752.

# A Event duration transformation proofs

This appendix provides some formal support to statements and explanations given in Section 4.

**Definition A.1.** Let $\mathbf{p}$ be a vector with $n$ elements where $n \geq 2$ and elements $p_{i=1} \dots p_n \in \mathbb{R}$. Let $\mathbf{d}$ be a vector with $m$ elements such that $\mathbf{d}_{ij} = p_i - p_j$ for all $i$ and $j = 1, \dots, n$ where $i > j$.

Equivalently $\mathbf{d}$ can be expressed as $\mathrm{vech}(\mathbf{p} \times \mathbf{1}_n^{\mathsf{T}} - \mathbf{1}_n \times \mathbf{p}^{\mathsf{T}})$. For ease of notation we use double indices to index each element of the vector $\mathbf{d}$. The indices $ij$ of the $k^{\text{th}}$ element of $\mathbf{d}$ are given by $d_{ij} = d_k = p_{i=\left\lfloor \frac{1}{2}+\sqrt{2k} \right\rfloor +1} - p_{j=k-C\left(\left\lfloor \frac{1+\sqrt{8k}}{2} \right\rfloor\right)}$ with $C[f(k)]$ being the binomial coefficient $\binom{f(k)}{2}$.

**Theorem A.1.** The number of elements in $\mathbf{d}$ is $m = n(n-1)/2$.

*Proof.* Because $d_{ij}$ must satisfy $i > j$ it can be thought of as all the elements below the subdiagonal of a $n \times n$ matrix of which there are $m = n(n-1)/2$. □

**Corollary A.1.1.** The sum of the number of elements in $\mathbf{p}$ and $\mathbf{d}$ is the number of edges on an $n+1$-polytope, or the number of edges in a complete graph with $n+1$ vertices.

*Proof.* The sum of the number of elements in $p$ and $d$ is $n + m = n + n(n-1)/2$ which is the number of edges on an $n+1$-polytope. $\square$

Consider the set $P$ of all possible choices of $n$ points in time.

**Definition A.2.** $P = \mathbb{R}^n$: Let $P = \mathbb{R}^n$ be the standard basis vector-space spanning all $\mathbf{p}$.

**Theorem A.2.** The linear transformation $T_{\mathbf{A}} : P \to D$ where $D = \mathbb{R}^{n(n-1)/2}$ and $\mathbf{A}_T\mathbf{p} = \mathbf{d}$ is given by

$$\underbrace{\left[\begin{array}{c|c|c} -\mathbf{I}_1 & \mathbf{1}_1 & \mathbf{0}_{1\times n-2} \\ -\mathbf{I}_2 & \mathbf{1}_2 & \mathbf{0}_{2\times n-3} \\ \vdots & \vdots & \vdots \\ -\mathbf{I}_{n-1} & \mathbf{1}_{n-1} & \left(\mathbf{0}_{n-1\times 0}\right) \end{array}\right]}_{\mathbf{A}_T} \times \underbrace{\begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}}_{\mathbf{p}} = \underbrace{\begin{bmatrix} d_1 \\ \vdots \\ d_{n(n-1)/2} \end{bmatrix}}_{\mathbf{d}}.$$

Given *any* $n$ points in time it is always possible to calculate the durations between any pair of these points. This calculation can be understood as a linear transformation of the *point-space* $P$ to the *duration-space* $D$.[6]

example: CPD $\mathbf{p}$ yields TAL $\mathbf{d}$

*Proof.* $A_T$ is a simple consequence of the definition of $\mathbf{d}$. $\square$

**Corollary A.2.1.** There exists no linear map $T_{\mathbf{A}}^{-1} : D \to P$.

Given only a set of $m$ durations it is impossible to identify a single set of $n$ points in time marking the endpoints of the durations, e.g. given only age one can not derive birth cohort or period.

*Proof.* $T_{\mathbf{A}}^{-1}$ exists if and only if $\mathbf{A}_T^{-1}$ exists. $\mathbf{A}_T$ has no inverse since the columns of $\mathbf{A}_T$ are always linearly dependent on each other, i.e. the last element in each row of $\mathbf{A}_T$ is always the negation of the sum of all the other row elements. Therefore $T_{\mathbf{A}}^{-1}$ does not exist. $\square$

**Theorem A.3.** *(Work in progess)* Something that states that you can transform $P^n \leftrightarrow M^n$ where the basis vectors of $M$ are a mixture of point dimensions and duration dimensions.

For example APT yields APCTDL. APT is a mixture of 1 element of $\mathbf{p}$ and 2 elements of $\mathbf{d}$. This is different from pure $P \to D$
TR: I think this theorem can be proved using graph theory as in the corollary below.

**Definition A.3.** $\mathbf{g}$ Let $\mathbf{g} = \{p_{i=1,\dots,n}, d_{k=1,\dots,m}\}$

**Theorem A.4.** (in progress) There are $b = (n+1)^{(n-1)}$ many ways to choose $n$ elements out of $\mathbf{g}$ whoose linear combination yields the remaining $m$ elements of $\mathbf{g}$.

[6] $\mathbf{I}_n$ is the $n \times n$ identity matrix; $\mathbf{1}_n$ is a vector with $n$ elements where each element is 1; $\mathbf{0}_{m\times n}$ is a $m \times n$ matrix where each element is 0. For notational convenience we allow for a matrix with 0 columns, written $\left(\mathbf{0}_{n\times 0}\right)$

*Proof.* This is a case of Cayley's formula (Cayley 1889), a result from graph theory which gives the number of possible trees on $k = n + 1$ vertices, $k^{(k-2)}$. In our case, the fully conected graph $\mathbf{K}$ with edges defined by the elements of $\mathbf{g}$ according to the third column of Table 3, is complete. By Cayley's fomula, the number of minimal spanning trees on $\mathbf{K}$ is equal to $k^{(k-2)}$. Four different proofs of this result are given in Aigner et al. (2010). The key is to realize that a minimal spanning tree (MST) on a complete graph will have $k - 1$ edges, connected to each other and all $k$ vertices. As such, the remaining possible edges are linear combinations of any given MST. □

**Corollary A.4.1.** Each set of $n$ elements from $\mathbf{g}$, $\mathbf{b}'$ whose linear transformation yields the remaining $m$ elements of $\mathbf{g}$ includes at least one element of $\mathbf{p}$.

*Proof.* One of the vertices of $\mathbf{K}$, say the $k^{\text{th}}$ vertex, is connected only to edges labelled by the elements of $\mathbf{p}$. Since a spanning tree of $\mathbf{K}$ must connect to this vertex to fully connect $\mathbf{K}$, all $b$ valid spanning trees must contain at least one edge labelled by an element of $\mathbf{p}$. □

TR: is the statement in this proof something that itself also needs to be proved?– that one vertex connects only to $\mathbf{p}$ If that statement needs to be proved, is it a lemma? I think this one takes care of the "mixing" theorem A.3 that is a work in progress above.

**Definition A.4.** $\mathbf{G}$ Let's define $\mathbf{G}$ as the identity implied by $\mathbf{g}$ whose graph is $\mathbf{K}$.

**Theorem A.5.** An identity $\mathbf{G}$ implies a total of $\binom{n+1}{3}$ triad sub-identities.

*Proof.* Any set of three vertices from $\mathbf{K}$ forms a complete subgraph, and any complete subgraph implies an identity between its labelled edges. $\mathbf{K}$ has $n+1$ vertices, and therefore there are $\binom{n+1}{3}$ ways to select three vertices from $\mathbf{K}$, hence $\mathbf{G}$ implies the same number of triad subidentities. □

**Corollary A.5.1.** Of the $\binom{n+1}{3}$ triad identies implied by $\mathbf{G}$, $\binom{n}{2}$ contain exactly two events and one duration.

*Proof.* This follows by noting that the $n$ event-labelled edges in $\mathbf{K}$ connect to a single vertex. Selecting any two of these $n$ event-labelled edges implies a tree on three vertices, whose full connection implies a triad identity composed of the two event edges and one duration edge defined as the time-difference of the former two. There are $\binom{n}{2}$ ways to select two of of the $n$ event edges. □

**Corollary A.5.2.** For $n \geq 3$, of the $\binom{n+1}{3}$ triad identies implied by $\mathbf{G}$, $\binom{n+1}{3} - \binom{n}{2} = \binom{n}{3}$ are composed of exactly three durations.

*Proof.* This is equivalent to deleting the vertex $k^{\text{th}}$ from $\mathbf{K}$, the vertex that connects only to event-labelled edges, which is constructed following the middle column of graphs from Table 3 with vertex labels ignored. This graph has $n$ total vertices, and any set of 3 vertices implies an identity between its three labelled edges, which in this case by definition can only consist of durations. □

For example, in the demographic time identity there are $n = 3$ event measures. Thus of the $\binom{4}{3} = 4$ triad identities implied $\binom{3}{3} = 1$ of these identities consist in durations only (TAL). Notice that the measures T and A change over the lifecourse of an individual, whereas their sum L is fixed.

**Definition A.5.** $\mathbf{d}_t$ For $\mathbf{p}$ that include period itself, let $\mathbf{d}_t$ be the set of duration time measures that change over the life course and $\mathbf{d}_f$ consist in those durations that are fixed attributed of an individual. By definition, $\mathbf{d}_t \cup \mathbf{d}_f = \mathbf{d}$.

awkward: how to say that $\mathbf{p}$ includes P? For example, we have P, but Lexis marriage does not, so this corolarry is only relevant for us and Brinks here.

**Corollary A.5.3.** For $n \geq 4$ and For $\mathbf{p}$ that include period itself, of the $d' = \binom{n}{3}$ triad identies whose edges are labelled only by the elements of $\mathbf{d}$, $d^t = \binom{d'}{2}$ of these identities consist in exactly two elements of $\mathbf{d}_t$ and one element of $\mathbf{d}_f$, while $d' - d^t$ of the duration-only triad identities consist in relationships between three elements of $\mathbf{d}_f$.

*Proof.* $n-2$ of the edges in $K$ are labelled with the elements of $\mathbf{d}_t$, and these all connect to the same vertex. There are therefore $\binom{n-2}{2}$ ways to form triad identities with them. The third element of each of these identities cannot connect to the same vertex, and so must be a member of $\mathbf{d}_f$. (this is almost done, needs a bit more logic) $\square$

**Theorem A.6.** In general, the number of subidentities of size $h$ in $\mathbf{G}$ is equal to $\binom{n+1}{n+1-h} \quad \forall\, h \leq n$.

*Proof.* Vertex deletion on a complete graph results in a complete subgraph. Therefore, the number of possible complete subgraphs with $h$ vertices is a function of the number of ways that $n+1-h$ vertices can be deleted from $\mathbf{K}$, which is $\binom{n+1}{n+1-h}$. The labelled edges of each possible complete subgraph defined in this way represent subidentities. $\square$

For example, from the tetrahedral graph in Figure 5b, we may delete the vertex that joins the edges labelled A, T, and P, which in effect deletes these edges, leaving us with the CDL identity.