



Hierarchical multistate models from population data: an application to parity statuses

Robert Schoen

Population Research Institute, Pennsylvania State University, University Park, PA, United States of America

ABSTRACT

Hierarchical models are characterized by having N living states connected by $N - 1$ rates of transfer. Demographic measures for such models can be calculated directly from counts of the number of persons in each state at two nearby points in time. Exploiting the ability of population stocks to determine the flows in hierarchical models expands the range of demographic analysis. The value of such analyses is illustrated by an application to childbearing, where the states of interest reflect the number of children a woman has born. Using Census data on the distribution of women by age and parity, a parity status life table for US Women, 2005–2010, is constructed. That analysis shows that nearly a quarter of American women are likely to remain childless, with a 0–3 child pattern replacing the 2–4 child pattern of the past.

Subjects Public Health, Women’s Health, Statistics

Keywords Multistate models, Parity status life tables, Childlessness, Hierarchical models, Sequential cross-sections, Polytrees

INTRODUCTION

Multistate population models are widely used in demographic analyses, as they can reflect any number of states and any pattern of movements between those states. The first multistate analysis is apparently due to *DuPasquier (1912)*, but multistate analyses received little attention until the 1970s (e.g., *Rogers, 1975*). In the 1980s, their nature and significance to demography was well established (*Land & Rogers, 1982; Schoen, 1988*). In general, a model with N living states has N^2 possible transfers: $N(N - 1)$ between living states, and N from each state to the “dead” state.

Population data on the number of persons in each state at the beginning and end of an age/time interval can provide only N equations, $N - 1$ for transfers between living states and 1 for the overall rate of death. Various procedures have been advanced to estimate all of the elements of the matrix that transforms the population at the beginning of an interval into the population at the end of the interval.

The leading technique is iterative proportional fitting (IPF), also known as the Deming-Stephan procedure, bi-proportional filling, and the RAS method (*Bishop, Fienberg & Holland, 1975; Willekens, 1982*). IPF yields the solution that maximizes entropy, i.e., it finds the transfers that can occur in the greatest number of ways. *Schoen & Jonsson (2003)* advanced an alternative, more demographic, approach based on estimating changes in

Submitted 14 June 2016
Accepted 6 September 2016
Published 29 September 2016

Corresponding author
Robert Schoen, rschoen309@att.net

Academic editor
Andrew Noymer

Additional Information and
Declarations can be found on
page 10

DOI 10.7717/peerj.2535

© Copyright
2016 Schoen

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

the attractiveness of model states. More recently, [Schoen \(2016\)](#) presented a Quadratic Estimation of Rates of Transfer (QERT) procedure, based on the assumption that the products of selected pairs of rates can be assumed constant. All three approaches assume a pattern that is believed to characterize the interstate movements, and thus produce estimated rather than calculated values.

This article focuses on “hierarchical” multistate models, that is on N state models with only $N - 1$ possible transfers connecting the living states and 1 unknown risk of death. In hierarchical models, a direct algebraic solution for the transfer matrix is possible, without the need to assume any pattern of interstate movements. We first discuss the parallels between hierarchical models and graphs, then present techniques for using sequential population data to calculate rates of transfer and transition probabilities, and proceed to apply the approach to calculate a Parity Status Life Table for US Women, 2005–10.

THE NATURE OF THE HIERARCHICAL MODEL

The N living states of a hierarchical model are linked by $N - 1$ occurrence/exposure rates of the form m_{ij} , where i is the origin state and j is the destination state. The dead state is indicated by the symbol δ . For every living state there is at least one connection to another living state, though all states are not necessarily reachable from any given state. For example, consider a model with two living states, A and B. With interstate transfer rate m_{AB} , it is possible to go from state A to state B, but not from state B to state A. The hierarchical model greatly restricts the range of possible interstate movements. It is a special case of the general multistate model, but one where the rates over an interval can be found from the populations at the beginning and end of the interval.

Graph theory can provide some useful nomenclature and context. The living states of a multistate model correspond to the vertices or nodes of a graph, and the rates connecting them are the edges or connecting lines of the graph. Hierarchical multistate models are connected graphs with N vertices and $N - 1$ edges, and hence constitute a “tree” ([Chartrand & Zhang, 2012](#)). Since the rates take persons from one specified state to another, the tree is “oriented” or, using the term coined by [Rebane & Pearl \(1987\)](#), a “polytree.”

Graph theorists have explored the number of distinct polytrees, or distinct multistate configurations, in models with N states (or vertices). When $N = 2$, there is only one form, with the single transfer rate m_{12} . When $N = 3$, there are 3 forms, which can be described by the rate pairs (m_{12}, m_{23}) , (m_{12}, m_{13}) , and (m_{13}, m_{23}) . There is no general rule for the number of forms, but when $N = 4$ there are 8 forms, when $N = 5$ there are 27 forms, and when $N = 6$ there are 91 forms ([Chartrand & Zhang, 2012](#)). Polytrees thus encompass a substantial number of distinct hierarchical multistate models.

Polytrees have demographic significance because transfers in every multistate model that is a polytree can be fully determined from knowledge of the population, by state, at the beginning and end of every interval of observation. Such multistate models can be used to analyze of a wide range of demographic behaviors, including parity progression, educational attainment, career trajectories, and the spread of infections. A procedure for carrying out the calculations needed to produce a hierarchical multistate life table from cross-sectional data is presented in the next sections.

FINDING MULTISTATE RATES FROM BEGINNING AND ENDING POPULATIONS

With N states, the beginning and ending populations support N equations that describe the flows (movements) between model states. Those N equations can be used to determine the $N - 1$ rates of transfer between those living states, and an overall rate of transfer to death. Each distinct hierarchical model (polytree) has a different set of N flow equations, but all those sets of equations are readily solvable because their N equations are independent.

For example, consider the case where $N = 3$ and the polytree form is (m_{12}, m_{13}) . The flow equations can then be written

$$\begin{aligned} P_1(x+n, n, t+n) &= P_1(x, n, t) - PP_1(x, n, t)[m_8(x, n, t) + m_{12}(x, n, t) + m_{13}(x, n, t)] \\ P_2(x+n, n, t+n) &= P_2(x, n, t) - PP_2(x, n, t)m_8(x, n, t) + PP_1(x, n, t)m_{12}(x, n, t) \\ P_3(x+n, n, t+n) &= P_3(x, n, t) - PP_3(x, n, t)m_8(x, n, t) + PP_1(x, n, t)m_{13}(x, n, t) \end{aligned} \quad (1)$$

where $P_j(x, n, t)$ denotes the number of persons in state j between the ages of x and $x+n$ at time t , and $PP_j(x, n, t)$ is the number of person-years lived in state j between times t and $t+n$ by persons aged x to $x+n$ at time t . The person-year values can be found from the beginning and ending population values in a number of ways (Schoen, 1988, Chap. 4). Here we use the linear relationship

$$PP_j(x, n, t) = (n/2)[P_j(x, n, t) + P_j(x+n, n, t+n)] \quad (2)$$

that assumes that the $P(x, n, t)$ function is linear between ages x and $x+n$. Even with 5-year age intervals, that linear assumption has been found to be satisfactory for most demographic work.

From Eqs. (1), straightforward algebra yields the solutions

$$\begin{aligned} m_8(x, n, t) &= \frac{P_1(x, n, t) + P_2(x, n, t) + P_3(x, n, t) - P_1(x+n, n, t+n) - P_2(x+n, n, t+n) - P_3(x+n, n, t+n)}{PP_1(x, n, t) + PP_2(x, n, t) + PP_3(x, n, t)} \\ m_{12}(x, n, t) &= \frac{[PP_1 + PP_3][P_2(n) - P_2(0)] - PP_2[P_1(n) + P_3(n) - P_1(0) - P_3(0)]}{PP_1(x, n, t)[PP_1(x, n, t) + PP_2(x, n, t) + PP_3(x, n, t)]} \\ m_{13}(x, n, t) &= \frac{[PP_1 + PP_2][P_3(n) - P_3(0)] - PP_3[P_1(n) + P_2(n) - P_1(0) - P_2(0)]}{PP_1(x, n, t)[PP_1(x, n, t) + PP_2(x, n, t) + PP_3(x, n, t)]} \end{aligned} \quad (3)$$

where the age and time identifiers have been dropped from the PP_j functions in the numerators and the age and time identifiers of the beginning and ending populations are simply denoted by (0) and (n), respectively, to simplify the presentation. The overall death rate is just the beginning number of persons minus the ending number, divided by the total number of person-years lived in the interval. The expressions for the transfer rates are a bit more complicated, but represent a weighted difference in state sizes over the interval, divided by the origin state number of person-years. Note that there is no assumption restricting a person to only one transfer during an interval (cf. Schoen, 1988, Ch. 4).

When the transfer rates take persons only from state 1 to state 2, state 2 to state 3, etc., the model can be termed “strictly” hierarchical. The flow equations of an N state strictly

hierarchical model are

$$\begin{aligned}
 P_1(x+n, n, t+n) &= P_1(x, n, t) - PP_1(x, n, t)[m_\delta(x, n, t) + m_{12}(x, n, t)] \\
 P_j(x+n, n, t+n) &= P_j(x, n, t) - PP_j(x, n, t)[m_\delta(x, n, t) + m_{j,j+1}(x, n, t)] \\
 &\quad + PP_{j-1}(x, n, t)m_{j-1,j}(x, n, t) \\
 P_N(x+n, n, t+n) &= P_N(x, n, t) - PP_N(x, n, t)m_\delta(x, n, t) \\
 &\quad + PP_{N-1}(x, n, t)m_{N-1,N}(x, n, t)
 \end{aligned} \tag{4}$$

where $1 < j < N$. Those flow equations yield the death and transfer rates

$$m_\delta(x, n, t) = \frac{\sum_i [P_i(x, n, t) - P_i(x+n, n, t+n)]}{\sum_i [PP_i(x, n, t)]} \tag{5a}$$

$$m_{j,j+1}(x, n, t) = \frac{\sum_i [P_i(x, n, t) - P_i(x+n, n, t+n) - PP_i(x, n, t)m_\delta(x, n, t)]}{PP_j(x, n, t)} \tag{5b}$$

where the sum in Eq. (5a) goes from 1 to N , and the sum in Eq. (5b) goes from 1 to j . The overall death rate, $m_\delta(x, n, t)$, is again the difference between the beginning and ending populations divided by the total number of person-years lived in the interval. The transfer rate from state j to state $j+1$ is the difference between the beginning and ending populations through state j , less the losses to death through state j , divided by the number of person-years lived in state j during the interval.

In general, the flow equations that describe the specific polytree form (or the structure of the desired multistate model), plus the initial and final population numbers in each state, can yield all of the transfer rates of a hierarchical multistate model. With the rates known, all model functions can readily be found (cf. Schoen, 1988). The only assumptions involved are that the population data are accurate, the population is closed, there is the same rate of death (or attrition) from every state, and the linear calculation assumption is appropriate. All of those assumptions can be relaxed with additional information. For example, if mortality is known, the interstate transfer rates can be found from cross-sectional surveys of the beginning and ending populations.

FINDING PROBABILITIES OF TRANSFER

The basic problem in life table construction is how to go from rates of transfer to probabilities of transfer. In multistate models, including hierarchical models, the problem is best approached by arraying the transfer rates in a matrix. With N living states, the $N \times N$ rate matrix can be written

$$\mathbf{M}(x, n, t) = \begin{bmatrix} \sum_j m_{1j} & -m_{12} & -m_{13} & \cdots & -m_{1N} \\ -m_{21} & \sum_j m_{2j} & -m_{23} & \cdots & -m_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -m_{N1} & -m_{N2} & -m_{N3} & \cdots & \sum_j m_{Nj} \end{bmatrix} \tag{6}$$

where the interval identifiers have been dropped from the m 's and the sums over j in the diagonal elements exclude m_{jj} and include the rate to the dead state. In hierarchical models, only N transfer rates, including m_δ , are nonzero.

There are a number of ways to calculate the array of transition probabilities from \mathbf{M} . One is to assume that the underlying risks of transition are constant over the interval and find the probabilities by exponentiating \mathbf{M} (cf. [Schoen, 2006b](#), Chap.1). Here, we will continue to use the linear assumption underlying [Eq. \(2\)](#), and employ the general linear solution of [Rogers & Ledent \(1976, Eq. \(2\)\)](#), as written in [Schoen \(2006b, Chap. 1\)](#). We then have

$$\mathbf{\Pi}(\mathbf{x}, \mathbf{n}, \mathbf{t}) = [\mathbf{I} - (n/2)\mathbf{M}(\mathbf{x}, \mathbf{n}, \mathbf{t})][\mathbf{I} + (n/2)\mathbf{M}(\mathbf{x}, \mathbf{n}, \mathbf{t})]^{-1} \quad (7)$$

where \mathbf{I} is the $N \times N$ identity matrix. Transition probability matrix $\mathbf{\Pi}(\mathbf{x}, \mathbf{n}, \mathbf{t})$ is an $N \times N$ matrix whose ij th element is $\pi_{ij}(x, n, t)$, the probability that a person in state i at age x and time t will be in state j exactly n years later. In hierarchical models, $\mathbf{\Pi}$ can be written as an upper triangular matrix, that is a matrix where all of the elements below the main diagonal are zero. That can be accomplished by numbering the states so that every nonzero m_{ij} has $i < j$.

The age-specific transition probability matrices for a period allow the construction of a multistate life table for that period. That life table provides, among other functions, the number of persons in each state at every age. Those life table survivorship values follow from the projection relationship

$$\ell^T(\mathbf{x} + \mathbf{n}) = \ell^T(\mathbf{x})\mathbf{\Pi}(\mathbf{x}, \mathbf{n}, \mathbf{t}) \quad (8)$$

where the j th element of the N element survivorship vector $\ell(\mathbf{x})$ is $\ell_j(x)$, the number of persons in the life table cohort who are in state j at exact age x . The superscript \mathbf{T} indicates that column vector $\ell(\mathbf{x})$ is transposed into a row vector. The initial (radix) vector, $\ell(\mathbf{0})$, is assumed to be known. The remaining multistate life table functions follow in the usual manner (cf. [Schoen, 1988, Chap. 4](#)).

The calculation that provides the transition probabilities from observed adjacent population values proceeds from the observed populations to the rates of transfer and then to the transition probabilities. It is thus rate centered, not probability based. Rates are preferable because they directly reflect the underlying behavior, i.e., the risk of movement from state i to state j , while probabilities only reflect that behavior in the context of other prevailing risks. In the hierarchical model, there are only $N - 1$ nonzero rates of transfer between living states while, because of multiple moves within an interval, there can be $N(N - 1)/2$ nonzero probabilities of transfer. That larger (for $N > 2$) number of probabilities cannot be determined from adjacent population data alone, unless the underlying rate structure is hierarchical. The proposed rate based approach accommodates multiple moves with an interval via the underlying Markov assumption that the risk of transfer depends only on current age and state.

AN APPLICATION TO PARITY STATUS LIFE TABLES

The procedure described in the previous two sections can find the rates and probabilities of any hierarchical multistate model from initial and final population data alone. To illustrate the procedure numerically, let us consider an important demographic application: the parity status model.

Table 1 Percentages of women by age and parity, United States 2005 and 2010.

Age	Parity							Total
	0	1	2	3	4	5	6	
YEAR 2005								
10–14	100	0	0	0	0	0	0	100
15–19	93.35	4.55	1.45	0.55	0.10	0	0	100
20–24	68.65	18.55	9.25	2.65	0.70	0.10	0.1	100
25–29	44.90	23.15	20.05	8.30	2.65	0.55	0.4	100
30–34	26.90	21.35	29.05	15.55	4.90	1.35	0.9	100
35–39	19.25	18.20	34.85	18.40	6.15	1.75	1.4	100
YEAR 2010								
15–19	94.7	4.4	0.6	0.3	0	0	0	100
20–24	70.5	18.1	9.0	2.0	0.3	0.1	0	100
25–29	47.6	22.7	18.7	7.8	2.3	0.6	0.3	100
30–34	29.7	19.2	29.1	14.3	5.2	1.5	1.0	100
35–39	19.7	18.5	32.7	19.7	5.9	2.1	1.4	100
40–44	18.8	18.5	33.3	19.1	6.8	1.9	1.6	100

Notes.

Note: percentages for 2005 are the average of those in 2004 and 2006.

Source: [Census.gov/hhes/fertility/data/cps/2004.html](https://www.census.gov/hhes/fertility/data/cps/2004.html), [Census.gov/hhes/fertility/data/cps/2006.html](https://www.census.gov/hhes/fertility/data/cps/2006.html) and [Census.gov/hhes/fertility/data/cps/2010.html](https://www.census.gov/hhes/fertility/data/cps/2010.html), downloaded 9/11/2014.

The demographic significance of parity

Parity, the number of live births a woman has had, is a meaningful demographic indicator. Its importance as a fertility measure was recognized in the 1950s (e.g., [Henry, 1953](#); [Whelpton, 1954](#)), and by the 1960s it was incorporated into population projections (cf. [Akers, 1965](#)). Parity is associated with subsequent fertility as well as with many other social, economic, and demographic behaviors. Childlessness has been a particular focus of interest, and has been seen as having strong connections to core social values ([Bulcroft & Teachman, 2004](#); [Koropecykj-Cox & Pendell, 2007](#); [Veevers, 1980](#)).

Data on births by age and parity are now frequently published by national statistical agencies (e.g., [Heuser, 1976](#)). Life tables recognizing parity status go back at least to [Chiang & Van Den Berg \(1982\)](#). Previous research has shown that parity changes were important in understanding fertility fluctuations (e.g., [Ryder, 1969](#)), and that the ultimate parity distributions of American women have varied greatly. A parity status life table based on 1970 fertility rates implied that 12% of women would remain childless, while 25% would have 4 or more children. Under 1995 rates, the life table proportion childless increased to 15%, while the proportion with 4 or more children dropped to 14% ([Schoen, 2006a](#), Table 1). Recent population data on levels of childlessness among American women aged 40–44 indicate that 18% were childless in 1995 and 15% in 2002 ([Abma & Martinez, 2006](#)).

US parity distributions and the rates they imply

The age distribution of American women, by number of children ever born, is available for even numbered years from the Census Bureau's Current Population Survey, accessible

Table 2 Rates of movement to successive parities, by age, US women 2005–10.

Age	Rate of movement from parity j to parity $j + 1$ ($m_{j,j+1}$)					
	m_{01}	m_{12}	m_{23}	m_{34}	m_{45}	m_{56}
12.5–17.4	.01089	.08182	.20000	0	0	0
17.5–22.4	.05620	.16777	.07464	.07843	.30000	.30000 ^a
22.5–27.4	.07243	.16388	.10662	.08804	.09333	.11429
27.5–32.4	.08150	.18087	.08220	.07257	.07898	.11707
32.5–37.4	.06180	.10088	.04146	.02553	.04630	.05797
37.5–42.4	.00473	.00163	.00998	.01067	.01081	.02192

Notes.

^aAdjusted value; calculated rate of 0.8 based on very few exposures.

Source: calculated from figures in Table 1 as described in text.

online at <http://www.census.gov/hhes/fertility/data/cps/>. Table 1 shows the proportions of women by age and parity for 2005 and 2010, the former year found as the average of the values for 2004 and 2006. Entries have been rounded so that all rows sum to exactly 100. There were very few women at parities higher than 6, so 6 was taken to be the highest parity possible. In 2010, 18.8% of women aged 40–44 were childless, while 1.6% attained parity 6.

The figures in Table 1, along with the assumption of no mortality at the reproductive ages, is sufficient to allow the construction of a parity status life table for the 2005–2010 interval. The model is strictly hierarchical, as women of parity j can only move to parity $j + 1$. The transfer rates of the model follow from Eq. (5), and are shown in Table 2. The age intervals shown are offset by 2.5 years from those in Table 1, as $P_j(x, n, t)$ was taken to be the number of persons at exact age $x + n/2$. Fertility rates are assumed to be zero for women past age group 40–44 in 2010. That assumption, while dictated by the available data, is not problematic because the rates at ages 37.5–42.5 are already quite small, and the rates at higher ages are even smaller.

The rates of transfer to parities 1 and 2 reach their maximum at ages 27.5–32.5, while the rates to parities 3 and 4 reach their maximum at ages 22.5–27.5. At young ages, rates to the highest parities are based on few observations, and tend to be unstable (e.g., the rate to parity 6 at ages 17.5–22.5). However, those rates are applied to the very small numbers of persons at those high parities, and hence have little impact on the life table parity distributions.

Calculating the transition probabilities and life table survivorship

Using Eq. (7), the rates of transfer can be used to calculate transition probability matrices for every age. Because the model is strictly hierarchical, the elements of the Π matrices are patterned, and can be written out in scalar form. The j th diagonal element has the basic life table survivorship form

$$\pi_{jj} = [2 - n m_{j-1,j}] / [2 + n m_{j-1,j}]. \quad (9)$$

The element in the j th column ($1 < j < N$) of the first row is given by

$$\pi_{1j} = [4n^{j-1} \prod_{i=1}^{j-1} m_{i-1,i}] / [\prod_{i=1}^j (2 + n m_{i-1,i})]. \quad (10)$$

In general, $\pi_{i+k,j+k}$, the element in the $(i+k)$ th row and $(j+k)$ th column ($i+k < j+k < N$), equals π_{ij} with all rate indexes increased by k . For example, in the parity status life table with $n = 5$, the probability of going from parity 2 to parity 4 is given by

$$\pi_{24} = \pi_{13}[\text{all rate indexes increased by 1}]$$

or

$$\pi_{24} = [100 m_{12}m_{23}]/\{[2 + 5 m_{12}][2 + 5 m_{23}][2 + 5 m_{34}]\}. \quad (11)$$

The elements of $\mathbf{\Pi}$ in the N th column are different because that state has no movements (decrements) to another state. Equation (7) yields

$$\pi_{1N} = [2n^{N-1} \prod_{i=1}^{N-1} m_{i-1,i}]/[\prod_{i=1}^{N-1} (2 + n m_{i-1,i})] \quad (12)$$

which differs from Eq. (10) in that the initial factor is 2 rather than 4, and that the product in the denominator goes to $N - 1$ instead of N . The element in the i th row ($i > 1$) and N th column is similarly the expression given in Eq. (11), but with an initial factor of 2 and a denominator product ending with the $N - 1$ st term.

Under the linear assumption, the expressions for the probabilities of moving more than one state in strictly hierarchical models can be interpreted in terms of successive risks of transfer. The probability of moving from state $j - 1$ to state j is $[n m_{j-1,j}]/[1 + (n/2)m_{j-1,j}]$. The linear assumption implies that moves occur uniformly over the interval or, on average, at the midpoint, $n/2$. The movers are then exposed to the risk of moving from state j to state $j + 1$ according to rate $m_{j,j+1}$ over half the interval. The successive factors in Eq. (10) reflect the likelihood of those additional moves, each with a shorter time exposure to risk.

The parity status life table

At birth, all women are at parity zero, so the parity distribution at initial age 12.5 has one person at parity zero and zero persons at all higher parities. The life table parity distributions at successively higher ages are found using projection Eq. (8).

Table 3 shows the parity distributions at ages 12.5 through 42.5 for US Women based on the data for 2005–2010. The proportion at parity zero steadily decreases, while the proportions at parities 3 and above steadily increase. The proportion at parity 1 reaches a peak at age 27.5, while the proportion at parity 2 peaks at age 37.5. At age 42.5, 23.4% of women are still at parity zero, while only 8.6% have 4 or more children. The US proportion childless has continued to rise, and is approaching one out of every four women. In contrast, large families of 4 or more children characterize barely one woman in twelve.

Many parity-specific analyses examine parity progression ratios (PPRs). The PPR from parity j to parity $j + 1$ is the proportion of persons who attain parity j who go on to attain parity $j + 1$. PPRs can be calculated from the ℓ_j functions. For example, at age 42.5, the PPR from parity 2 to parity 3 is the sum of the ℓ_j values for parity 3 and above divided by the sum of the ℓ_j values for parity 2 and above, or $.264/.570 = .463$. That PPR is considerably higher than the PPR from parity 3 to parity 4, which is $.326$.

Table 3 Parity distributions and cumulated fertility, by age, parity status life table for US women 2005–10.

Age	Proportion of women at parity indicated							Cumulated fertility to age indicated
	0	1	2	3	4	5	6	
12.5	1.	0	0	0	0	0	0	0
17.5	.947	.044	.006	.003	0	0	0	.06
22.5	.714	.182	.084	.016	.002	.001	.001	.42
27.5	.495	.232	.183	.069	.016	.004	.002	.90
32.5	.327	.203	.283	.129	.041	.011	.006	1.41
37.5	.240	.191	.320	.172	.050	.017	.010	1.69
42.5	.234	.195	.306	.178	.056	.018	.012	1.73

Notes.

Cumulated fertility is the sum of the births to the life table cohort (of 1 woman) up to the given age. Parity 6 is assumed to be the highest parity possible, and no fertility is assumed after age 42.5. Thus cumulated fertility to age 42.5 represents the Parity Status Total Fertility Rate.

Source: calculated from Table 2 as described in text.

The Parity Specific Total Fertility Rate (PSTFR) of the life table cohort can be found from the parity distribution of women at the end of childbearing, here taken to be age 42.5. Algebraically,

$$\text{PSTFR} = \sum_{j=1}^N (j-1)\ell_j(42.5) \quad (13)$$

where $\sum \ell_j(42.5) = 1$ and state 1 is parity 0. The PSTFR differs from the conventional TFR because it is calculated from age-parity-specific fertility rates, rather than the usual age-specific fertility rates, and thus takes the parity composition of the hypothetical cohort into account. Table 3 shows a PSTFR, or cumulative fertility to age 42.5, of 1.73, less than two children per woman.

The life table proportion childless, 0.234, is substantially larger than 0.188, the proportion childless among American women in 2010. That observed proportion reflects childlessness in one actual cohort of women. In contrast, the life table proportion indicates that a greater fraction of women would never have a child under the age-parity-specific fertility rates prevailing during 2005–2010. The initiation of childbearing in the observed cohort followed a different pattern from that in the 2005–2010 period.

SUMMARY AND CONCLUSIONS

Many demographic phenomena can be represented by hierarchical models, that is by models with N living states that have $N - 1$ rates of movement between them. In such models, the rates of transfer between living states and to death can be found from data on the number of persons in each state at the beginning and end of a given interval. With the rates of transfer provided by Eq. (5), the probabilities of transfer can be found from Eq. (7) and used to produce a multistate life table. Exploiting the ability of population counts to infer rates of transfer in hierarchical contexts broadens the scope of demographic analysis.

The potential of the approach is illustrated by an application to parity status. A parity status life table for US Women, 2005–2010, is constructed from survey data on the fraction of women, by age, in each parity status in 2004, 2006, and 2010. The results show that the US proportion childless is continuing to increase, with nearly a quarter of women aged 42.5 remaining at parity zero. The 2–4 child pattern of the past has been replaced by a 0–3 child pattern.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The author received no funding for this work.

Competing Interests

The author declares there are no competing interests.

Author Contributions

- Robert Schoen conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.

Data Availability

The following information was supplied regarding data availability:

The raw data has been supplied as [Supplementary Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.2535#supplemental-information>.

REFERENCES

- Abma JC, Martinez GM. 2006.** Childlessness among older women in the United States: trends and profiles. *Journal of Marriage and Family* **68**:1045–1056 DOI [10.1111/j.1741-3737.2006.00312.x](https://doi.org/10.1111/j.1741-3737.2006.00312.x).
- Akers DS. 1965.** Cohort fertility versus parity progression as methods of projecting births. *Demography* **2**:414–428 DOI [10.2307/2060128](https://doi.org/10.2307/2060128).
- Bishop YM, Fienberg SE, Holland PE. 1975.** *Discrete multivariate analysis: theory and practice*. Cambridge: MIT Press.
- Bulcroft R, Teachman J. 2004.** Ambiguous constructions: development of a childless or child-free life course. In: *Handbook of contemporary families*. New York: Sage Publications, 116–135.
- Chartrand G, Zhang P. 2012.** *A first course in graph theory*. Mineola: Dover.
- Chiang CL, Van Den Berg BJ. 1982.** A fertility table for the analysis of human reproduction. *Mathematical Biosciences* **62**:237–251 DOI [10.1016/0025-5564\(82\)90085-2](https://doi.org/10.1016/0025-5564(82)90085-2).

- DuPasquier LG. 1912.** Mathematische theorie der invaliditätsversicherung. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 7:1–7.
- Henry L. 1953.** Fécondité des mariages: Nouvelle méthode de mesure. In: *INED/PUF Travaux et documents*. cahier No. 16. Paris: Presses Universitaires De France.
- Heuser RL. 1976.** *Fertility tables for birth cohorts by color: United States 1917–73*. Rockville: US Dept of Health , Education, and Welfare.
- Koropeckyj-Cox T, Pendell G. 2007.** The gender gap in attitudes about childlessness in the United States. *Journal of Marriage and Family* 69:899–915 DOI 10.1111/j.1741-3737.2007.00420.x.
- Land KC, Rogers A (eds.) 1982.** *Multidimensional mathematical demography*. New York: Academic Press.
- Rebane G, Pearl J. 1987.** The recovery of causal poly-trees from statistical data. In: *Proceedings of the third annual conference on uncertainty in artificial intelligence*. Seattle, WA.
- Rogers A. 1975.** *Introduction to multiregional mathematical demography*. New York: Wiley.
- Rogers A, Ledent J. 1976.** Increment-decrement life tables: a comment. *Demography* 13:287–290 DOI 10.2307/2060807.
- Ryder NB. 1969.** The emergence of a modern fertility pattern: United States 1917–66. In: Behrman SJ, Corsa L, Freedman R, eds. *Fertility and family planning: a world view*. Ann Arbor: University of Michigan Press, 99–126.
- Schoen R. 1988.** *Modeling multigroup populations*. New York: Plenum.
- Schoen R. 2006a.** Insights from parity status life tables for the 20th century US. *Social Science Research* 35:29–39 DOI 10.1016/j.ssresearch.2004.06.002.
- Schoen R. 2006b.** *Dynamic population models*. Dordrecht: Springer.
- Schoen R. 2016.** Multistate transfer rate estimation from adjacent populations. *Population Research and Policy Review* 35:217–240.
- Schoen R, Jonsson SH. 2003.** Estimating multistate transition rates from population distributions. *Demographic Research* 9(29 August):1–24 DOI 10.4054/DemRes.2003.9.1.
- Veevers JE. 1980.** *Childless by choice*. Toronto: Butterworth.
- Whelpton PK. 1954.** *Cohort fertility, native white women in the United States*. Princeton: Princeton University Press.
- Willekens FJ. 1982.** Multidimensional population analysis with incomplete data. In: Land KC, Rogers A, eds. *Multidimensional mathematical demography*. New York: Academic, 43–111.