

# Stochastic models of telomere shortening

Peter Olofsson, Marek Kimmel \*

*Department of Statistics, Rice University, P.O. Box 1892, Houston, TX 77251, USA*

Received 3 February 1998; received in revised form 14 September 1998; accepted 30 November 1998

---

## Abstract

Shortening of chromosome ends, known as telomeres, is one of the supposed mechanisms of cellular aging and death. We provide a probabilistic analysis of the process of loss of telomere ends. The first work concerned with that issue is the paper by Levy et al. [*J. Molec. Biol.* 225 (1992) 951–960]. Their deterministic model reproduced the observed frequencies of viable cells in the in vitro experiments. Arino et al. [*J. Theor. Biol.* 177 (1995) 45–57] reformulated the model of Levy et al. (1992) in the terms of branching processes with denumerable type space. In the present paper, the mathematical results of Arino et al. (1995) are extended to the case in which cell death is present, in cells with telomeres above and below the critical threshold of length, generally with differing probabilities. Both exact and asymptotic results are provided, as well as a discussion of biological relevance of the results. © 1999 Elsevier Science Inc. All rights reserved.

*Keywords:* Telomeres; Cell aging; Reducible multitype branching process; Asymptotic behavior

---

## 1. Introduction

Shortening of chromosome ends known as telomeres, is one of the supposed mechanisms of cellular aging and death. As discussed in the review by Harley [1], it has been known that telomeres play an important role in chromosome structure and function. Under normal circumstances, aberrations involving

---

\* Corresponding author. Tel.: +1-713 285 5255; fax: +1-713 285 5476; e-mail: kimmel@rice.edu

telomeres are extremely rare. On the other hand, broken chromosomes lacking telomeres are highly unstable, producing translocations, fusions and other aberrations. Thus telomere deletions could have severe consequences on cell functions. Also, incomplete replication of DNA at the ends of linear chromosomes is predicted from the known biochemical characteristics of DNA replication, unless distinct mechanisms exist for replication of ends of DNA molecules. It is possible that somatic cells lack these mechanisms (used by immortal cells to completely replicate telomeres) and as a natural consequence suffer loss of some telomeric DNA at every round of cell division.

This loss of telomeres is proposed by some cell biologists as an explanation for the finite proliferative capacity of cell lines. Some cells, known as immortal cells, have arrested telomere shortening due to the presence of enzyme telomerase, which replaces the lost nucleotides upon each division. This includes some cancer cells, which continue to divide with much shorter telomeres and thus undergo many more cell divisions than normal cells. The role of chromosomal telomere loss is a fundamental problem in cell biology.

Experimental evidence, gathered mainly from primary fibroblast cell cultures, suggests that cells possess a genetic mitotic clock which influences their division capacity. In the terminal passages of the cell cultures, there is a fairly dramatic increase in the frequency of cells with chromosomal aberrations, notably telomere fusions. These karyotypic changes may result from and not initiate the decline of the culture. However, it is possible that chromosomal aberrations and the senescent decline are both initiated by the same events.

Two plausible molecular mechanisms for these events have been proposed. Olovnikov [2] suggested that cells lose a small amount of DNA following each round of replication due to the inability of DNA polymerase to fully replicate chromosome ends (telomeres), and that eventually, after a certain number of divisions, a critical deletion causes cell death. Holliday [3] proposed that DNA replication was accompanied by progressive methylation to repetitive DNA, which would lead to cell senescence after a certain number of divisions.

Olovnikov's [2] 'end-replication' problem is the inability of DNA polymerase to fully replicate the ends of a linear DNA molecule. DNA consists of strands that are antiparallel and DNA polymerases require a primer to initiate unidirectional synthesis. Thus, following replication each duplex would be shortened on the 5' end of the daughter DNA strand. This type of telomere loss generates a 3' single stranded overhang which, if not degraded, is converted to a double stranded deletion in the subsequent generation.

The aim of the present paper is to provide foundations for a rigorous probabilistic analysis of the process of loss of telomere ends. The first work concerned with that issue is the paper by Levy et al. [4]. We base our considerations on ideas developed in that paper, the authors of which developed a simple mathematical model of the process. The model was essentially deterministic and did not account directly for the observed variability of cell pro-

liferation. Nevertheless, it reproduced the observed frequencies of viable cells in the in vitro experiments.

Arino et al. [5] reformulated the model of Levy et al. [4] in the terms of branching processes with denumerable type space. This approach made possible to include variability in cell cycle times. They provided mathematical results regarding dynamics of the models, in particular their essentially polynomial asymptotics. Also, they fitted the models to data from two different sources, using an approach differing from that of Levy et al. [4]. Their main conclusion was that the end-replication hypothesis provided an explanation for the cell aging process which was quantitatively consistent with the data.

In the present paper, the mathematical results of Arino et al. [5] are extended to the case in which cell death is present, in cells with telomeres above and below the critical threshold of length, generally with differing probabilities. Both exact and asymptotic results are provided, as well as a discussion of biological relevance of the results.

## 2. Model of the loss of telomeres

The model describes shortening of telomeres by incomplete replication. In accordance with the models by Levy et al. [4] and by Arino et al. [5] we assume that telomeres are the components of the terminal restriction fragments (TRFs) detectable by a molecular probe. The TRF consists of *telomeres* which are the terminal TTAGGG or TTAGGG-like DNA repeats, and of non-terminal *subtelomeric* non-TTAGGG-like DNA repeats.

The two uses of the model are prediction of: (1) the expected telomere length and (2) of the fraction of viable cells, in aging cell populations. For these purposes, it is first necessary to describe the dynamics of telomere loss from a single chromosome. For simplicity, we proceed as if the process of telomere loss ended when all the telomere *deletion units*, each containing possibly more than a single TTAGGG-like repeat, are lost. The same mathematics applies to telomere loss until a particular *checkpoint* is encountered.

A chromosome is an entity with a centromere, while a chromatid is a double helix composed of two single strands of DNA. In  $G_1$  phase of the cell cycle, before DNA replication, a chromosome is composed of one chromatid, while in  $G_2$  and M phases, after DNA replication, a chromosome is composed of two chromatids. Levy et al. [4] described telomere loss in terms of what happens to single DNA strands in  $G_1$  (cf. Fig. 2 in that paper). We follow that description.

Our Fig. 1 depicts the scheme of deletion and segregation of telomere sequences on the ends of two DNA strands constituting the chromatid in  $G_1$ , following Levy et al. [4]. These rules can be summarized as follows:

- Each chromatid is composed of two strands named *upper* or  $5' \rightarrow 3'$ , and *lower* or  $3' \rightarrow 5'$ , each of which has two ends named *left* and *right*. The num-

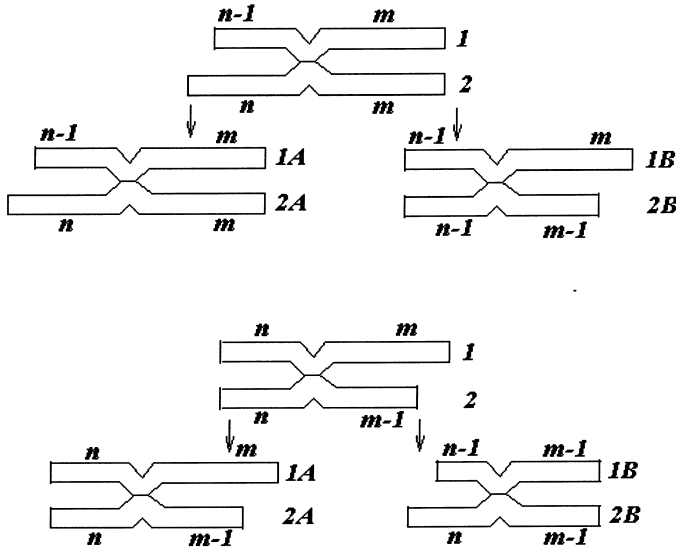


Fig. 1. Transition rules for deletion and segregation of telomere ends on a chromosome in the  $G_1$  phase of the cell cycle. DNA strands 1 and 2 replicate and segregate into different daughter cells A and B, resulting in chromatids (1A, 2A) and (1B, 2B), respectively. Due to the end-replication problem, one DNA strand on each of the newly created chromatids contains additional deletion at its right or left end, depending on its orientation and presence of the deletion on the corresponding strand of the mother chromatid. For additional explanations, cf. Fig. 1 in Ref. [4].

bers of telomeric deletion units on both ends of both strands are listed as the quadruples of the form  $(a, b; c, d)$ , where  $a$  and  $c$  correspond to the left and right ends of the upper strand, while  $b$  and  $d$  correspond to the left and right ends of the lower strand. The only admissible combinations of  $a, b, c$  and  $d$  are of the form  $(n - 1, n; m, m)$  or  $(n, n; m, m - 1)$ .

- Cells containing chromatids described by the quadruple  $(n - 1, n; m, m)$  give birth to two progeny containing chromatids of the types  $(n - 1, n; m, m)$  and  $(n - 1, n - 1; m, m - 1)$ , respectively. This transition rule as well as an analogous rule for the other admissible type are depicted symbolically below. Let us note that one progeny is always of the same type as the parent cell, while the other is missing two sequences, each on a different end of a different strand:

$$(n - 1, n; m, m) \begin{cases} \rightarrow & (n - 1, n; m, m), \\ \rightarrow & (n - 1, n - 1; m, m - 1), \end{cases}$$

$$(n, n; m, m - 1) \begin{cases} \rightarrow & (n, n; m, m - 1), \\ \rightarrow & (n - 1, n; m - 1, m - 1), \end{cases}$$

- The process ends when the telomere ends become short enough. Without loss of generality, we assume that cells of the types  $(n - 1, n; 0, 0)$  and

$(0, 0; m, m - 1)$  have a single progeny of the type identical to that of the parent cell, i.e.:

$$(n - 1, n; 0, 0) \rightarrow (n - 1, n; 0, 0),$$

$$(0, 0; m, m - 1) \rightarrow (0, 0; m, m - 1).$$

**Remark 1.** Let us notice that each non-identity transition of our process corresponds to the loss of *two* deletion units. This is important when fitting the model to data.

If we renumber states in such way that index  $k = 0, 1, \dots$ , is equal to the sum of numbers of deletion units on the left ends of the upper and lower strand, and index  $l = 0, 1, \dots$ , is equal to the sum of numbers of deletion units on the right ends of the upper and lower strand:

$$k = \begin{cases} 2n & \text{if } (n, n; m, m - 1) \text{ occurs,} \\ \text{or} & \\ 2n - 1 & \text{if } (n - 1, n; m, m) \text{ occurs,} \end{cases} \quad (1)$$

$$l = \begin{cases} 2m & \text{if } (n - 1, n; m, m) \text{ occurs,} \\ \text{or} & \\ 2m - 1 & \text{if } (n, n; m, m - 1) \text{ occurs,} \end{cases} \quad (2)$$

then the admissible transitions become

$$(k, l) \begin{cases} \rightarrow (k, l), \\ \rightarrow (k - 1, l - 1), \end{cases} \quad (3)$$

$$(k, 0) \rightarrow (k, 0), \quad (4)$$

$$(0, l) \rightarrow (0, l). \quad (5)$$

In the array  $(k, l)$ , where  $k$  and  $l$  are non-negative integers, the admissible transitions belong to disjoint paths which can be numbered by  $k - l$  (path number assuming values from  $-\infty$  through  $\infty$ ). Each of these paths can be treated separately. The state number within each path can be taken as  $i = \min(k, l)$ . Biologically, it is the number of deletion units on the shorter, and therefore limiting, end. Now the transitions have the form

$$i \begin{cases} \rightarrow i, \\ \rightarrow i - 1, \end{cases} \quad (6)$$

$$0 \rightarrow 0. \quad (7)$$

### 3. Main results

We start by briefly describing the mathematical model and main result of Arino et al. [5]. The evolution of the process is as follows: lifelengths of cells are

independent, identically distributed random variables with the common density  $f(t) = \alpha e^{-\alpha t}$ , i.e. exponentially distributed with mean  $1/\alpha$ . At the moment of death, a cell of type  $i$  splits into two cells, one of type  $i$  and one of type  $i - 1$ , for  $i \geq 1$ . A cell of type 0 simply replaces itself at the moment of death, or put in a different way, 0-type cells are immortal.

Now assume that the process is started with one cell of type  $i \geq 1$ . We are interested in the expected number of cells of type  $i - k$ ,  $k \leq i$  at time  $t$ , denoted by  $M_{i,i-k}(t)$ . It turns out that it is useful in the analysis to consider also the expected number of cells of type  $i - k$  in the  $n$ th generation, denoted by  $m_{i,i-k}^{(n)}$  which in the model in Ref. [5] turns out to equal  $nk$ .

The main result is that, under these assumptions,

$$M_{i,i-k}(t) = \frac{\alpha^k t^k}{k!} \quad (8)$$

for  $k \leq i$ . It is further shown that, if lifelengths are not exponential, the same formula holds asymptotically if the lifelengths are bounded with mean  $1/\alpha$ .

We consider two extensions of these results. First, and most important, we consider the exponential case and introduce possible cell death into the model, i.e. the possibility that cells die without dividing. This gives qualitatively different results in terms of the composition of the cell population. From a theoretical point of view, we note that the methods we use are different from those in Ref. [5]. There, analytical methods are used whereas we use combinatorial and probabilistic methods which turn out to work in greater generality.

It should be pointed out here that both Theorems 1 and 2 in the next section may be proved by using general results for Markovian branching processes. These can be found for instance in Ref. [6]. We give alternative proofs that in this particular situation are more intuitive and also in principle extend beyond the Markovian case.

Second, we explore the asymptotics for the model in Ref. [5] for general lifelength distributions. The expression for  $M_{i,i-k}(t)$  is now a certain infinite sum. It turns out that the asymptotics remain the same for arbitrary lifelength distributions, so for asymptotic properties, the assumption of exponential lifelengths is not crucial. For finite time properties, the infinite sum can be computed numerically for a given lifelength distribution. The asymptotic result is stated and proved in Appendix A.

### 3.1. Models with cell death

Let the ancestor be of type  $i$ . Recall  $M_{i,i-k}(t)$ , the expected number of  $i - k$ -type cells at time  $t$  and  $m_{i,i-k}^{(n)}$ , the expected number of  $i - k$ -type cells in the  $n$ th generation. The two relate in the following way:

$$M_{i,i-k}(t) = \sum_{n=k}^{\infty} m_{i,i-k}^{(n)} (G^{*n}(t) - G^{*(n+1)}(t)),$$

where  $G$  is the distribution function of the lifelength and  $G^{*n}$  its  $n$ -fold convolution, the distribution of the sum of  $n$  i.i.d. lifelengths. To understand this formula, note that at time  $t$ , cells from any generation may be present. Since there are  $m_{i,i-k}^{(n)}$  individuals in the  $n$ th generation and each of these is alive at time  $t$  with probability  $G^{*n}(t) - G^{*(n+1)}(t)$  (born before  $t$  but not yet dead at  $t$ ), the expected number of cells from the  $n$ th generation present at time  $t$  is simply  $m_{i,i-k}^{(n)}(G^{*n}(t) - G^{*(n+1)}(t))$ . Summing over all the generations then gives us  $M_{i,i-k}(t)$ .

The identity may be formally derived as the unique solution to a certain renewal equation, for a more general version of this, see Ref. [7].

In particular, if lifelengths are exponential ( $\alpha$ ), it is well known that

$$G^{*n}(t) = e^{-\alpha t} \sum_{k=0}^{n-1} \frac{(\alpha t)^k}{k!},$$

a gamma distribution with parameter  $n$  and  $\alpha$  and hence

$$M_{i,i-k}(t) = e^{-\alpha t} \sum_{n=k}^{\infty} m_{i,i-k}^{(n)} \frac{(\alpha t)^n}{n!}. \tag{9}$$

If there is no cell death,  $m_{i,i-k}^{(n)} = \binom{n}{k}$  and hence

$$M_{i,i-k}(t) = e^{-\alpha t} \sum_{n=k}^{\infty} \binom{n}{k} \frac{(\alpha t)^n}{n!} = \frac{(\alpha t)^k}{k!} e^{-\alpha t} \sum_{n=k}^{\infty} \frac{(\alpha t)^{n-k}}{(n-k)!} = \frac{(\alpha t)^k}{k!}$$

as in Arino et al. [5].

**Hypothesis 1.** 0-type cells survive with probability  $p$  and die with probability  $1 - p$ .

Clearly,  $M_{i,i-k}(t)$  will be as above as long as  $k < i$ ; when  $k = i$  we have an asymptotic result:

**Theorem 1.** Under Hypothesis 1, if cell lifelengths are exponential with parameter  $\alpha$ ,

$$M_{k,j}(t) = \frac{(\alpha t)^{k-j}}{(k-j)!}, \quad j = 1, \dots, k$$

and

$$M_{k,0}(t) \sim \frac{1}{1-p} \frac{(\alpha t)^{k-1}}{(k-1)!},$$

as  $t \rightarrow \infty$ .

*Note.* The lifelength  $L_0$  of a 0-type is exponential ( $\alpha(1-p)$ ). To see this, note that 0-types die with probability  $(1-p)$  at the points in a Poisson process with intensity  $\alpha$ . With  $N(t)$  denoting the number of points up to time  $t$  in this process we thus obtain

$$\begin{aligned} P(L_0 > t) &= \sum_{k=0}^{\infty} p^k P[N(t) = k] = \sum_{k=0}^{\infty} p^k e^{-\alpha t} \frac{(\alpha t)^k}{k!} \\ &= e^{-\alpha(1-p)t}. \end{aligned}$$

Therefore, by introducing cell death for 0-type cells, we are actually modelling the situation where non-zero cells have lifelengths that are exponential with mean  $1/\alpha$  and zero cells have lifelengths that are exponential with another mean  $1/\tau$ . By choosing  $p$  suitably we can achieve this for any  $\tau$ . Hence we have a model where there are no longer any immortal cells, and two different death rates.

*Note.* The factor  $1/(1-p)$  is the mean in a geometric distribution with success probability  $1-p$ .

**Proof.** See Appendix A.

In the model with immortal cells, the 0-cells tend to dominate since their growth rate is the fastest. In the model with cell death, the 1-cells have the same growth rate so asymptotically they will dominate together with the 0-cells in the proportions 1 to  $1/(1-p)$ . This is already a qualitative difference between the two models. Next we introduce the possibility that also non-zero cells die without reproducing.

**Hypothesis 2.** Assume that a cell survives with probability  $p$  if it is of type 0 and with probability  $q$  otherwise.

The result is now shown in Theorem 2.

**Theorem 2.** Under Hypothesis 2, if the lifelengths are exponential with parameter  $\alpha$

$$M_{k,j}(t) = \frac{(\alpha q t)^{k-j}}{(k-j)!} e^{-\alpha(1-q)t}, \quad j = 1, \dots, k.$$

(i) If  $p = q$ , then

$$M_{k,0}(t) = \frac{(\alpha q t)^k}{k!} e^{-\alpha(1-q)t}.$$

(ii) If  $p < q$ , then

$$M_{k,0}(t) \sim \frac{q}{q-p} \frac{(\alpha q t)^{k-1}}{(k-1)!} e^{-\alpha(1-q)t}.$$



(iii) If  $p > q$ , then

$$M_{k,0}(t) \sim \left( \frac{q}{p - q} \right)^k e^{-\alpha(1-p)t}.$$

**Proof.** See Appendix A.

Note that the polynomial growth is now asymptotically killed by an exponential decay factor, determined by the larger of the two survival probabilities  $p$  and  $q$ . A natural generalization would be to let a cell of type  $i$  survive with probability  $p_i$  but for the applications we are primarily concerned with in this paper, this is not of much use. This, and other generalizations will be attempted elsewhere.

### 3.2. Arbitrary lifelength distributions

Consider the case when the lifelengths have the common distribution function  $G$  and there is no cell death. We know that

$$M_{k,j}(t) = \sum_{n=k-j}^{\infty} \binom{n}{k-j} (1 - G) * G^{*n}(t).$$

The asymptotic behaviour of  $M_{k,j}(t)$  is given in Theorem 3.

**Theorem 3.** *Assume that the lifelengths are not identically 0. Then, as  $t \rightarrow \infty$ ,*

$$M_{k,j}(t) \sim \frac{t^{k-j}}{\mu^{k-j}(k-j)!},$$

where  $\mu < \infty$  is the mean lifelength.

The above result is an extension of the result in Ref. [5]. Their result, as represented in our expression (8), is obtained as a special case from Theorem 3 by setting  $\mu = \alpha^{-1}$ .

## 4. Computational results

In telomere shortening experiments, the quantities observed experimentally are: (i) the fraction  $F(t)$  of proliferating cells, and (ii) the mean length  $E(t)$  of telomere endings, in culture. These two quantities are given by the following expressions:

$$F(t) = \frac{\sum_{j=1}^k M_{k,j}(t)}{\sum_{j=0}^k M_{k,j}(t)}, \tag{10}$$

$$E(t) = \frac{\sum_{j=1}^k jM_{k,j}(t)}{\sum_{j=0}^k M_{k,j}(t)}. \quad (11)$$

Another potentially observable quantity is the total number of cells in culture,  $\sum_{j=0}^k M_{k,j}(t)$ . However, this latter is rarely accounted for, as cell cultures proceed through series of passages. Let us notice that according to Theorem 3, the ratio of the number of proliferating to non-proliferating cells exhibits different limit properties depending on the relationship between  $p$  and  $q$ . As can be directly verified, as  $t \rightarrow \infty$ ,

$$\frac{\sum_{j=1}^k jM_{k,j}(t)}{M_{k,0}(t)} = O\left(\frac{1}{t}\right); \quad 0 < p = q \leq 1,$$

$$\frac{\sum_{j=1}^k jM_{k,j}(t)}{M_{k,0}(t)} \rightarrow \frac{q-p}{q}; \quad 0 < p < q \leq 1,$$

$$\frac{\sum_{j=1}^k jM_{k,j}(t)}{M_{k,0}(t)} = O[t^{k-1}e^{-\alpha(p-q)t}]; \quad 0 < q < p \leq 1,$$

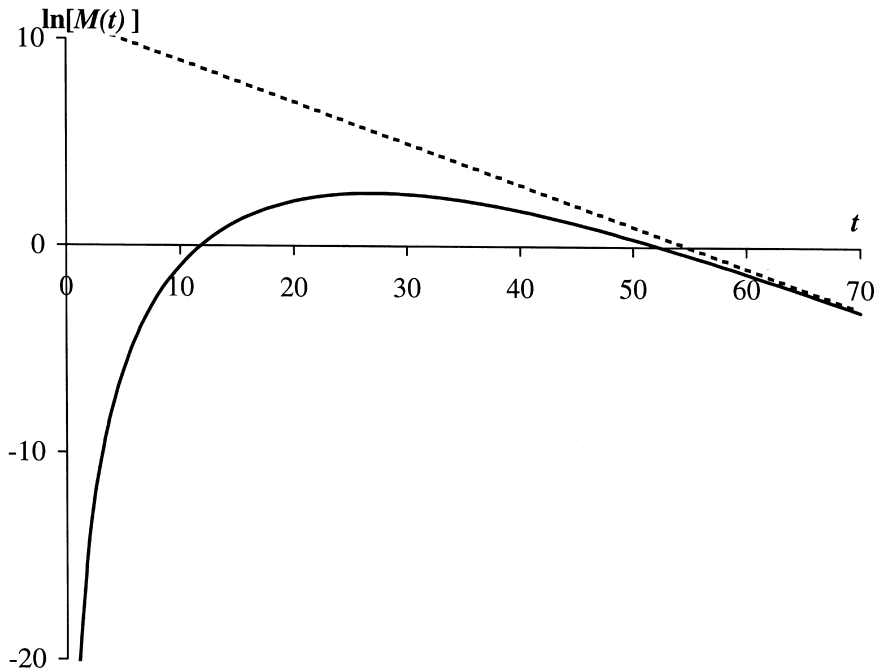


Fig. 2. Convergence of  $M_{k,0}(t)$  to its asymptotics for a combination of parameters:  $p = 0.8$ ,  $q = 0.6$ ,  $\alpha = 1$ , and  $k = 10$ . *Continuous line*: exact value; *dotted line*: asymptotics.

while the total number of cells eventually tends to 0, except for cases  $p = 1$  or  $q = 1$ . Particularly interesting is the case  $0 < p < q \leq 1$ , in which

$$F(t) \rightarrow \left( \frac{q-p}{q} \right) \left[ \left( \frac{q-p}{q} \right) + 1 \right]^{-1} = \frac{q-p}{2q-p} > 0,$$

as  $t \rightarrow \infty$ , i.e. the fraction of proliferating cells remains positive for arbitrarily long times. It can be demonstrated that in this case,  $E(\infty) = F(\infty)$ .

Fig. 2 depicts the convergence of  $M_{k,0}(t)$  to its asymptotics for a combination of parameters. Fig. 3 depicts the plots of  $F(t)$  and  $E(t)$  for another combination of parameters. Numerical computations were carried out using expressions (13) and (12) of the Appendix A.

## 5. Discussion

The purpose of this paper was to introduce refined models of telomere shortening in the form of a multitype branching process which allows for the possibility of cell death. We identified several interesting variants and provided exact computable expressions as well as asymptotic expressions for the expected values of the process. In one variant of the process,  $0 < p < q \leq 1$ , the fraction of proliferating cells remains positive for arbitrarily long times (although the total cell population eventually declines). This demonstrates that experimental cell lines may display apparent stability despite the fact that they do not become immortalized.

The current paper has a speculative character and the model plots are not intended to be compared to experimental data, at least not in a quantitative sense. However, let us notice that the *in vitro* telomere data are based on cell counts in excess of  $10^6$  cells (Levy et al. [4]) and therefore the estimates of Eqs. (10) and (11) can be reliably obtained as ratios of respective cell counts. Such estimates were employed by Levy et al. [4] and by Arino et al. [5].

The model considered in our paper is a particular multi-type Bellman–Harris process. Moreover, for exponential lifelengths, our results could be obtained using classical theorems (Mode [6]). However, from the point of view of possible applications, we need to know the constants arising in the theorems explicitly and it does not seem easier to compute the matrix powers that would have to be done using classical results. Our proofs, in these special cases, are closer to intuition. For the general lifelength distributions, we have found no results in the literature covering this particular case. The results in Mode's [6] book concern either the supercritical case (Perron–Frobenius roots greater than one) or the positively regular case.

We identified two sources of data that relate to our process. One is the data of Harley and Goldstein as cited by Levy et al. [4] in which fractions  $F(t)$  of proliferating cells were measured at different times after a clonal culture had

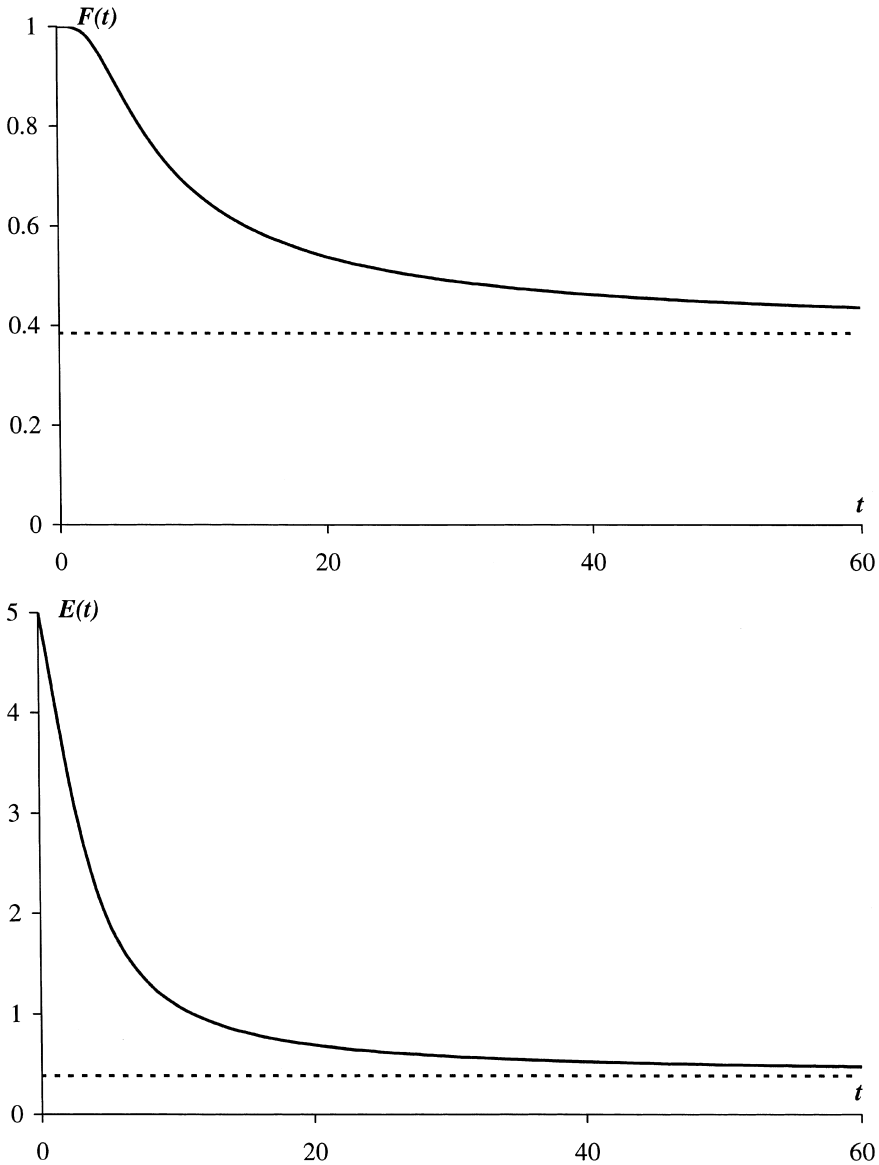


Fig. 3. Fraction of proliferating cells  $F(t)$  and mean length of telomeres  $E(t)$  for a combination of parameters:  $p = 0.3$ ,  $q = 0.8$ ,  $\alpha = 1$ , and  $k = 10$ . *Continuous lines:  $F(t)$  and  $E(t)$ ; dotted lines: asymptotic values.*

been established. These data have been used for modeling by Levy et al. [4] (see their Fig. 6). Another source is the paper by Counter et al. [8] which includes experimental data on the expected telomere lengths  $E(t)$ . The data were as-

sembled, plotted and reproduced by the mathematical model in the paper by Arino et al. [5]. Our plots in Fig. 3 qualitatively match the data.

The medical relevance of the process of telomere loss was already mentioned in the Introduction in the context of immortalization of cancer cells. Other important examples include shortening of telomeres in T lymphocytes in Type I diabetes (Jeanclos et al. [9]) and in some subsets of T lymphocytes in AIDS (Wolthers et al. [10]). These observations may reflect an increased turnover of lymphocytes in these diseases (i.e. more divisions of lymphocytes leading to shortening of telomeres) or, in case of Type I diabetes, inherited shorter telomeres contributing to the onset of disease. In healthy people, older individuals tend to have lymphocytes with shorter telomeres than younger individuals (Weng et al. [11]).

Our study does not include the case in which the enzyme telomerase, capable of restoring telomere endings, is activated (as presumably in immortal cells). Modeling telomerase action would lead to bidirectional branching random walks with different asymptotic properties.

The model of telomere loss can be formulated as a population dynamics model with the so called inherited property (Arino et al. [12]). A formulation of this latter model in the terms of branching processes can be found in Ref. [13].

### Acknowledgements

M.K. was supported by grants GM 58545 from the National Institutes of Health, and DMS 9409909 from the National Science Foundation, and by the Keck's Center for Computational Biology at the Rice University. M.K. did part of his research in April and May 1998, when he was a long-term visitor at the Gothenburg Stochastic Centre, supported by the Swedish Foundation for Strategic Research.

### Appendix A. Proofs

**Proof of Theorem 1.** First consider  $m_{k,0}^{(n)}$ . Pick a 0-type cell in the  $n$ th generation and follow its family tree backwards. This will consist of a number of 0-type cells and eventually a first 1-type cell. Assume that this occurs in generation  $j$ . The probability that this 1-type cell initiates a cell-line of 0's that still exists in the  $n$ th generation is  $p^{n-j-1}$ . Since there are  $m_{k,1}^{(j)}$  cells in the  $j$ th generation, the expected number of 0-type cells in the  $n$ th generation is

$$m_{k,0}^{(n)} = \sum_{j=k-1}^{n-1} m_{k,1}^{(j)} p^{n-j-1}$$

and since  $m_{k,1}^{(j)} = \binom{j}{k-1}$  we obtain

$$m_{k,0}^{(n)} = \sum_{j=k-1}^{n-1} \binom{j}{k-1} p^{n-j-1}.$$

By reversing the order of summation we may write this as

$$m_{k,0}^{(n)} = \sum_{i=0}^{n-k} \binom{n-i}{k-1} p^i$$

and since  $\binom{n-i}{k-1} \sim n^{k-1}/(k-1)!$  as  $n \rightarrow \infty$ ,

$$m_{k,0}^{(n)} \sim \frac{1}{(1-p)} \frac{n^{k-1}}{(k-1)!}.$$

From this we obtain the asymptotics of  $M_{k,0}(t)$ . First note that, for any  $N$ ,

$$\lim_{t \rightarrow \infty} M_{k,0}(t) = \lim_{t \rightarrow \infty} e^{-\alpha t} \sum_{n>N} m_{k,0}^{(n)} \frac{(\alpha t)^n}{n!}$$

and hence we can choose  $N$  large and replace  $m_{k,0}^{(n)}$  by  $1/(1-p)n^{k-1}/(k-1)!$  in Eq. (9). We obtain

$$\begin{aligned} M_{k,0}(t) &\sim \frac{1}{(1-p)} e^{-\alpha t} \frac{1}{(k-1)!} \sum_n n^{k-1} \frac{(\alpha t)^n}{n!} \\ &\sim \frac{1}{(1-p)} \frac{(\alpha t)^{k-1}}{(k-1)!} e^{-\alpha t} \sum_n \frac{(\alpha t)^{n-k+1}}{(n-k+1)!} \sim \frac{1}{(1-p)} \frac{(\alpha t)^{k-1}}{(k-1)!} \end{aligned}$$

as  $t \rightarrow \infty$ .

*Note.* To see that this is not an exact formula, let us compute  $M_{k,0}(t)$  explicitly for some values of  $k$ . For  $k=1$  we get

$$m_{1,0}^{(n)} = \frac{1}{1-p} (1-p^n)$$

and hence

$$M_{1,0}(t) = \frac{1}{1-p} e^{-\alpha t} (e^{\alpha t} - 1 - (e^{\alpha p t} - 1)) = \frac{1}{1-p} (1 - e^{-\alpha(1-p)t}).$$

And for  $k=2$ :

$$m_{2,0}^{(n)} = \sum_{j=1}^n j p^{n-j-1} = \frac{n(1-p) - p + p^{n+1}}{p(1-p)^2}$$

which yields

$$M_{2,0}(t) = \frac{1}{1-p} \left( \alpha t + \frac{1}{1-p} (e^{-\alpha(1-p)t} - 1) \right).$$

**Proof of Theorem 2.** First note that, if  $i > k$ ,  $m_{i,i-k}^{(n)} = \binom{n}{k} q^n$  and hence

$$M_{i,i-k}(t) = \frac{(\alpha q t)^k}{k!} e^{-\alpha t} \sum_{n=k}^{\infty} \frac{(\alpha q t)^{n-k}}{(n-k)!} = \frac{(\alpha q t)^k}{k!} e^{-\alpha(1-q)t}. \tag{A.1}$$

(i) In this case the formula above is valid also for  $i = k$ .

(ii) Here note that

$$m_{k,1}^{(j)} = \binom{j}{k-1} q^j$$

and that the probability that a 1-type cell in the  $j$ th generation initiates a line of 0-types that is still alive in the  $n$ th generation is  $q p^{n-j-1}$ . Hence

$$m_{k,0}^{(n)} = \sum_{j=k-1}^{n-1} \binom{j}{k-1} q^{j+1} p^{n-j-1} = q^n \sum_{j=k-1}^{n-1} \binom{j}{k-1} \left(\frac{p}{q}\right)^{n-j-1}.$$

Again reversing the order of summation yields

$$m_{k,0}^{(n)} = q^n \sum_{i=0}^{n-k} \binom{n-i}{k-1} \left(\frac{p}{q}\right)^i$$

and the asymptotics

$$m_{k,0}^{(n)} \sim q^n \frac{n^{k-1}}{(k-1)!} \frac{q}{q-p}.$$

As above we can replace  $m_{k,0}^{(n)}$  by its asymptotics in Eq. (9) to obtain

$$M_{k,0}(t) \sim \frac{q}{q-p} \frac{(\alpha q t)^{k-1}}{(k-1)!} e^{-\alpha t} \sum_n \frac{(\alpha q t)^{n-k+1}}{(n-k+1)!} \sim \frac{q}{q-p} \frac{(\alpha q t)^{k-1}}{(k-1)!} e^{-\alpha(1-q)t}.$$

(iii) First recall the identity

$$\sum_{n=k}^{\infty} \binom{n}{k} x^n = \frac{x^k}{(1-x)^{k+1}}$$

for  $0 \leq x \leq 1$ . Hence

$$\sum_{j=k-1}^{\infty} \binom{j}{k-1} \left(\frac{q}{p}\right)^j = \frac{p}{q} \left(\frac{q}{p-q}\right)^k$$

and

$$\begin{aligned} M_{k,0}(t) &= e^{-\alpha t} \sum_{n=k}^{\infty} \sum_{j=k-1}^{n-1} \binom{j}{k-1} q^{j+1} p^{n-j-1} \frac{(\alpha t)^n}{n!} \\ &= \frac{q}{p} \sum_{j=k-1}^{\infty} \binom{j}{k-1} \left(\frac{q}{p}\right)^j e^{-\alpha t} \sum_{n=j+1}^{\infty} \frac{(\alpha p t)^n}{n!} \\ &\sim \left(\frac{q}{p-q}\right)^k e^{-\alpha(1-p)t}. \end{aligned} \tag{A.2}$$

*Note.* If  $p = 1$  and  $k = 1$  we get  $M_{1,0}(t) \rightarrow q/(1 - q)$  which makes sense since the number of 0-types then will be equal to the number of surviving 1-types which is geometric including 0 with success probability  $1 - q$ . For  $k = 2$  we get  $(q/(1 - q))^2$  where the second factor is the mean number of 1-types stemming from a 2-type; this is also geometric including 0 and hence the mean is  $q/(1 - q)$ . For general  $k$  we get  $(q/(1 - q))^k$ .

**Proof of Theorem 3.** The proof is based on the following Tauberian theorem in Ref. [14].

**Theorem A.1.** *Let  $H$  be a non-decreasing function such that the Laplace–Stieltjes transform  $\hat{H}(s) = \int_0^\infty e^{-su}H(du) < \infty$  for  $s > 0$ . Then*

$$\hat{H}(s) \sim \frac{A}{s^r} \text{ as } s \rightarrow 0$$

*if and only if*

$$H(t) \sim \frac{At^r}{\Gamma(r + 1)} \text{ as } t \rightarrow \infty.$$

For the proof of Theorem 3, first let  $j = k - 1$ . Then

$$\begin{aligned} M_{k,k-1}(t) &= \sum_{n=1}^\infty n(1 - G) * G^{*n}(t) \\ &= \sum_{n=1}^\infty n(G^{*n}(t) - G^{*(n+1)}(t)) \\ &= \sum_{n=1}^\infty \sum_{i=1}^n (G^{*n}(t) - G^{*(n+1)}(t)) \\ &= \sum_{i=1}^\infty \sum_{n=i}^\infty (G^{*n}(t) - G^{*(n+1)}(t)) \\ &= \sum_{i=1}^\infty G^{*i}(t) \sim \frac{t}{\mu} \end{aligned}$$

as  $t \rightarrow \infty$ , by the renewal theorem in Ref. [15]. That this theorem applies can be realized directly: the  $k$ -type cells split according to a (delayed) renewal process and at each renewal an infinite line of  $(k - 1)$ -type cells is initiated since there is no death. Hence the number of  $(k - 1)$ -type cells at time  $t$  is exactly the number of such renewals up to time  $t$ .

In the absence of cell death, clearly  $M_{k,k-1}(t)$  is non-decreasing and



$$\begin{aligned} \hat{M}_{k,k-1}(s) &= \int_0^\infty e^{-su} M_{k,k-1}(\mathbf{d}u) = \sum_{n=1}^\infty \int_0^\infty e^{-su} G^{*n}(\mathbf{d}u) \\ &= \sum_{n=1}^\infty \hat{G}(s)^n = \frac{\hat{G}(s)}{1 - \hat{G}(s)} < \infty \end{aligned}$$

for  $s > 0$  unless  $G$  is defective (concentrated at 0). Hence, Theorem A.1 above applies and

$$\hat{M}_{k,k-1}(s) \sim \frac{1}{\mu s}$$

as  $s \rightarrow 0$ .

Now, consider  $M_{k,j}(t)$  for an arbitrary  $j$ . We obtain

$$\begin{aligned} \hat{M}_{k,j}(s) &= \int_0^\infty e^{-su} M_{k,j}(\mathbf{d}u) \\ &= \sum_{n=k-j}^\infty \binom{n}{k-j} \int_0^\infty e^{-su} (1 - G) * G^{*n}(\mathbf{d}u) \\ &= (1 - \hat{G}(s)) \sum_{n=k-j}^\infty \binom{n}{k-j} \hat{G}(s)^n \\ &= (1 - \hat{G}(s)) \frac{\hat{G}(s)^{k-j}}{(1 - \hat{G}(s))^{k-j+1}} \\ &= \hat{M}_1(s)^{k-j}. \end{aligned}$$

Hence

$$\widehat{M}_{k,j}(s) \sim \frac{1}{\mu^{k-j} s^{k-j}}$$

as  $s \rightarrow 0$  and applying Theorem A.1 again yields

$$M_{k,j}(t) \sim \frac{t^{k-j}}{\mu^{k-j} (k-j)!}$$

as  $t \rightarrow \infty$ .

### References

- [1] C.B. Harley, Telomere loss: mitotic clock or genetic time bomb?, *Mut. Res.* 256 (1991) 271.
- [2] A.M. Olovnikov, A theory of marginotomy, *J. Theor. Biol.* 41 (1973) 181.
- [3] R. Holliday, Growth and death of diploid and transformed human fibroblasts, *Fed. Proc.* 34 (1975) 51.

- [4] M.Z. Levy, R.C. Allsopp, A.B. Futcher, C.W. Greider, C.B. Harley, Telomere end-replication problem and cell aging, *J. Molec. Biol.* 225 (1992) 951.
- [5] O. Arino, M. Kimmel, G.F. Webb, Mathematical modeling of the loss of telomere sequences, *J. Theor. Biol.* 177 (1995) 45.
- [6] C. Mode, *Multitype Branching Processes*, Elsevier, New York, 1971.
- [7] P. Jagers, General branching processes as Markov fields, *Stoch. Proc. Appl.* 32 (1989) 183.
- [8] C.M. Counter, A.A. Avilion, C.E. LeFeuvre, N.G. Stewart, C.W. Greider, C.B. Harley, S. Bacchetti, Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity, *EMBO J.* 11 (1992) 1921.
- [9] E. Jeanclos, A. Krolewski, J. Skurnick, M. Kimura, H. Aviv, A. Aviv, Shortened telomere length in white blood cells of patients with IDDM, *Diabetes* 47 (1998) 482.
- [10] K.C. Wolthers, G.B.A. Wisman, S.A. Otto, A.-M. de Roda Husman, N. Schaft, F. de Wolf, J. Goudsmit, R.A. Coutinho, A.G.J. van der Zee, L. Meyaard, F. Miedema, T cell telomere length in HIV-1 infection: No evidence for increased CD4<sup>+</sup> T cell turnover, *Science* 274 (1996) 1543.
- [11] N.-P. Weng, B.L. Levine, C.H. June, R.J. Hodes, Human naive and memory T lymphocytes differ in telomeric length and replicative potential, *Proc. Nat. Acad. Sci. USA* 92 (1995) 11091.
- [12] O. Arino, E. Sanchez, G.F. Webb, Polynomial growth dynamics of telomere loss in a heterogeneous cell population, *Dynam. Contin. Discr. Impulsive Syst.* 3 (1997) 262.
- [13] P. Olofsson, A branching process model of telomere shortening, *Stochastic Models*, submitted.
- [14] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2, Wiley, New York, 1966.
- [15] S. Asmussen, *Applied Probability and Queues*, Wiley, New York, 1987.