CrossMark

Models for the red blood cell lifespan

Rajiv P. Shrestha[1] · Joseph Horowitz[2] · Christopher V. Hollot[3] · Michael J. Germain[4] ·
John A. Widness[5] · Donald M. Mock[6] · Peter Veng-Pedersen[7] · Yossi Chait[8]

Abstract The lifespan of red blood cells (RBCs) plays an important role in the study and interpretation of various clinical conditions. Yet, confusion about the meanings of fundamental terms related to cell survival and their quantification still exists in the literature. To address these issues, we started from a compartmental model of RBC populations based on an arbitrary *full lifespan* distribution, carefully defined the *residual lifespan*, *current age*, and *excess lifespan* of the RBC population, and then derived the distributions of these parameters. For a set of residual survival data from biotin-labeled RBCs, we fit models based on Weibull, gamma, and lognormal distributions, using nonlinear mixed effects modeling and parametric bootstrapping. From the estimated Weibull, gamma, and lognormal parameters we computed the respective population mean full lifespans (95 % confidence interval): 115.60 (109.17–121.66), 116.71 (110.81–122.51), and 116.79 (111.23–122.75) days together with the standard deviations of the full lifespans: 24.77 (20.82–28.81), 24.30 (20.53–28.33), and 24.19 (20.43–27.73). We then estimated the 95th percentiles of the lifespan distributions (a surrogate for the maximum lifespan): 153.95 (150.02–158.36), 159.51 (155.09–164.00), and 160.40 (156.00–165.58) days, the mean current ages (or the mean residual lifespans): 60.45 (58.18–62.85), 60.82 (58.77–63.33), and 57.26 (54.33–60.61) days, and the residual half-lives: 57.97 (54.96–60.90), 58.36 (55.45–61.26), and 58.40 (55.62–61.37) days, for the Weibull, gamma, and lognormal models respectively. Corresponding estimates were obtained for the individual subjects. The three models provide equally excellent goodness-of-fit, reliable estimation, and physiologically plausible values of the directly interpretable RBC survival parameters.

✉ Rajiv P. Shrestha
rajiv@octetresearch.com

[1] Octet Research Inc., 101 Arch St. Suite 1950, Boston, MA 02110, USA

[2] Department of Mathematics & Statistics, University of Massachusetts, Amherst, MA 01003, USA

[3] Department of Electrical & Computer Engineering, University of Massachusetts, Amherst, MA 01003, USA

[4] Renal and Transplant Associates of New England, Division of Nephrology, Baystate Medical Center, Tufts University School of Medicine, Boston, MA, USA

[5] Department of Pediatrics, College of Medicine, The University of Iowa, Iowa City, IA 52242, USA

[6] Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA

[7] Division of Pharmaceutics, College of Pharmacy, The University of Iowa, Iowa City, IA 52242, USA

[8] Department of Mechanical & Industrial Engineering, University of Massachusetts, Amherst, MA 01003, USA

Introduction

The survival of red blood cells (RBCs) has been studied for nearly a century [1] because of its importance in clinical medicine and translational research. To cite three examples, the mean RBC lifespan and production rate determine the steady-state hemoglobin (Hb) concentrations in both healthy and diseased individuals [2]; variation in mean RBC lifespan is sufficient to result in clinically important differences in HbA1c among diabetics with the same mean

blood glucose level [3]; and reduced mean RBC lifespan contributes to the anemia of chronic renal failure [4, 5]. Equally important is the role of the RBC lifespan in establishing criteria for stored RBCs for transfusion, elucidation of the physiology and pathophysiology of erythropoiesis, and the design of therapies for anemia and other hematological conditions [6–8].

Confusion and a lack of proper quantification of various aspects of RBC survival still exist in the literature. Before discussing these issues we describe the two approaches, direct and indirect, commonly used to study RBC survival and the types of RBC samples used in such studies [9]. We then provide a brief literature review.

Direct and indirect models of RBC survival

The *direct approach* [9] involves observing and quantifying the disappearance of a population of labeled RBCs from the circulation. Direct models fall primarily into two categories: (1) empirical models for simple curve fitting [3, 5, 8, 10–14], and (2) "phenomenological" models accounting for macroscopic phenomena such as random destruction of the cells, lifespan-based elimination of the cells, neocytolysis (selective hemolysis of young RBCs under conditions of RBC excess when acutely exposed to increased tissue oxygen content), etc., and method-specific phenomena such as radioactive decay of the label, elution, vesiculation, etc. in the case of labeling by ^{51}Cr [9, 10, 15–21].

Instead of the direct measurement of RBC survival, the *indirect approach* [9] relies on native Hb, glycosylated Hb, or other pharmacokinetic/pharmacodynamic (PK/PD) information over time to make inferences about the RBC lifespan [2, 9, 22–25]. Indirect models are lifespan-based compartmental PK/PD models (e.g., lifespan-based indirect response or LIDR models [26]), of which transit compartment (TC) models are a special case [27]. A series of transit compartments, each compartment with a PDF representing the lifespan distribution of the cells in that compartment [9, 27, 28] is a hallmark of TC models.

As the name indicates, LIDR models are "indirect" models. Direct models, in which the survival of labeled RBCs is tracked over time, are based on the residual survival function (SF), which is derived here as a consequence of the LIDR model.

Random and cohort samples

Direct RBC survival studies can be conducted in two ways depending on the type of sample of RBCs used [7]. A *random* (or *population*) sample consists of a mixture of RBCs of all ages, such as that obtained by a venous blood draw. In contrast, a *cohort* sample consists of RBCs that are all approximately the same age [7]. Random sample methods are more easily performed and thus more widely used in RBC survival studies [7, 14]. In such studies, RBCs in blood drawn from a donor are labeled and transfused into a recipient (possibly the same person), and their disappearance is followed [7]. The time at which the sample is collected is called the *index time*, and the corresponding RBC population (in the circulation plus in the sample) at that time is the *index population*.

For a random sample study, the sampling of blood and transfusion of labeled RBCs back into the subject should ideally happen at the index time and within a time interval insignificant in comparison to the shortest lifespan of all but a negligible fraction of the RBCs; further, the transfused RBCs should be distributed homogeneously in the circulation immediately. The concentration of labeled RBCs in the circulation at that instant would serve as the baseline concentration. In practice, the labeled RBC enrichment at day 1 is taken as the baseline to minimize artifacts from RBCs that are damaged and consequently removed during the first 24 h in circulation [29]. Blood samples are subsequently drawn at sufficient intervals after the labeled RBC transfusion to permit determination of RBC survival parameters. The ratio of the concentration of labeled RBCs in the later samples to the baseline concentration gives the survival data at each time. A plot of these measurements against time constitutes the random sample (residual) survival curve.

Labeling with radioactive ^{51}Cr is the current de facto gold standard for RBC survival studies [14], despite the fact that it exposes subjects to radiation and despite the analytical complications due to radioactive decay, elution of ^{51}Cr from cells, and loss of label by vesiculation [30]. Recently, Mock et al. [31] demonstrated that random sample survival data for normal adults obtained using nonradioactive biotin labeled RBCs (bioRBCs) can be used in RBC survival studies with results similar to those from ^{51}Cr. Rather than measuring the radioactivity of hemoglobin bound ^{51}Cr in the blood samples, which is inevitably confounded by radioactive decay, vesiculation, and elution, individual cells are enumerated by flow cytometry after separation from unlabeled RBCs based on fluorescence intensity due to binding of fluorescent-labeled streptavidin to the biotin label on the RBCs. Thus, this method is free from the problems of elution and vesiculation. Moreover, biotinylation at lower densities appears not to affect the RBC lifespan [14].

Brief literature survey

Mathematical descriptions of RBC survival are broadly based on the direct or indirect method described above. Direct models are based on the theory for transfused RBCs

developed early on [10, 32–34]. Some quantitative descriptions developed later focused on fitting a curve to the survival curve and quantifying the mean lifespan and half-life [5, 11, 15, 17, 18, 32]. Complicated phenomenological models were also used [10, 15–21, 30]. Some authors used simple linear or cubic curve-fitting of the survival curve to estimate the mean age at the time of labeling [3, 14]. Many studies use the maximum RBC lifespan (T_{max}), a term that is difficult to quantify [3, 11, 13] in a plausible manner.

A direct PK/PD model including RBC survivor functions of various complexity was presented by Uehlinger et al. [25]. A new lifespan based compartmental PK/PD model paved the way to an indirect and compartmental approach to quantify RBC/reticulocyte mean lifespan [35]. Earlier works based on this approach assumed that all cells in a compartment survived for a fixed time, which introduced a time-delay in the input/output models thus representing the lifespan of the cells in a compartment by a point distribution [2, 23, 26, 35–37]. The theory was expanded later to represent the RBC lifespan distribution by possibly time-varying continuous PDF [21, 22, 38–40]. These indirect methods do not involve direct measurement of random sample cell survival data. But Lledó-García et al. [9] presented a somewhat "hybrid" approach that used a TC model to estimate mean lifespan from a set of biotin-labeled random sample RBC survival data previously published by Cohen et al. [3].

There exist confusions in the literature primarily due to failure to distinguish between various survival functions and survival parameters and subtle disconnects in the theoretical relationships developed. Furthermore, some models have a large number of unknown parameters, which creates identifiability issues.

Objectives

In this paper, starting with a LIDR model based on an arbitrary lifespan distribution, we derive the relations between the residual lifespan distribution, which is the basis of observations in the direct method, and the full lifespan distribution; these are expressed through the survival functions corresponding to each distribution. Further, we derive survival functions for the current age distribution (distribution of ages in the index RBC population) and the excess lifespan distribution (distribution of remaining lifespan of cells in the index population that have survived beyond a specified time t_e past the index time). Other RBC survival descriptors (mean lifespan, mean current age, etc.) are then easily obtained.

We apply these results to analyze a set of bioRBC data [14] using the NLME framework. We consider three models,

based on the Weibull, gamma, and lognormal distributions, respectively, for the full RBC lifespan. The PDFs for these distributions are given in Appendix 1. These distributions are widely used in reliability theory [41–43], survival analysis [44], and studies of aging and lifespan for both mechanical and biological systems using a reliability theory approach [45–49]. Although the Weibull is widely used as a failure distribution [41], aside from the papers by Friese et al. [22, 39], ours is the only use of the Weibull as an RBC lifespan distribution to our knowledge. In RBC survival studies the gamma distribution arises in transit compartment models [27] with a fixed, specified integer-valued shape parameter, a restriction we avoid in this work.

Based on observations of the residual survival function provided by the experimental data, the estimated structural parameters of each model are used to compute the survival functions for the full lifespan, current age, and excess lifespan distributions. From these it is easy to estimate many RBC survival descriptors at both the population and individual levels; as examples we use the mean lifespan, the standard deviation of the lifespan distribution, τ_{95} (as a surrogate for maximum lifespan), mean current age, and residual half-life. The definitions of these terms are provided later in the text.

The process of RBC aging and the eventual death is an issue of special scientific and clinical interest [50–52]. The RBC lifespan distribution may provide information of clinical and research relevance in understanding the type of red cell elimination taking place in normal individuals and in those with specific diseases. For example, these models can potentially be used to determine whether elimination is primarily driven by an intrinsic lifespan-based mechanism or is the result of extrinsic factors acting independently of the RBC lifespan.

Post-transfusion survival of stored RBC has important implications in determination of the shelf life of RBCs [8, 53]. Excess lifespan of RBC can potentially be used to better quantify the post-transfusion survival of stored RBC.

A potential use of current age and excess lifespan distributions is to help identify morphological and/or physiological attributes of RBCs that correlate to the RBC age distribution, shedding light on the mechanism of RBC aging and eventual death in health and disease [54]. For example, it has been shown that the surface area and hemoglobin amount generally resemble lognormal distribution and the distribution changes in a time dependent manner as the membrane continues to be shed [55]. Time dependent changes in the density distribution of biotin-labeled sickle RBC have been studied in [56]. If we are able to accurately measure the current age or excess lifespan distributions based on some morphological or physiological attributes of RBCs, then it would potentially

allow the quantification of full RBC survival descriptors discussed in this paper with a single random sample of blood instead of multiple consecutive random samples over the period of up to 4 months as in the most commonly used residual lifespan data.

Mathematical model

The *birth time* of an RBC is the time when a reticulocyte transitions to a fully matured and hemoglobinized RBC after losing the ribosomes and mitochondria. By *death time* of RBC we mean the moment of elimination of the RBC from the circulation. At birth time $t = -u$, each RBC is assumed to be endowed with a *full lifespan* τ. Consider the index time $t = 0$ when the RBC sample is collected and labeled. The age of an RBC in the index population at this time is the *current age*, $\tau_c = u$, which is the amount of time already spent by it in circulation since its "birth." The *residual lifespan* of the RBC is the time τ_r that it remains in the circulation before its "death". Thus

$$\tau = \tau_c + \tau_r.$$

Given a fixed, specified time t_e after the index time, the remaining lifespan of the RBC is *excess lifespan* of the RBC, $\tau_e = \tau - t_e$. Not all RBCs are born with the same lifespans. The full lifespan and hence the residual lifespan, the current age, and the excess lifespan can be represented by respective continuous probability density functions (PDFs). In general the RBC survival distributions for different people will not be the same.

We assume that the RBCs in healthy individuals survive independently of each other. We assume the internal environment of the study subject is stable in the sense that the RBC lifespan distribution is time-invariant and that there is no abnormal blood loss or other significant intercurrent event (e.g., events leading to hemolysis).

Full lifespan distribution

The *(full) lifespan distribution* in the population of RBCs in a given individual is represented by a PDF $p(t)$, $t \geq 0$. The corresponding full SF is the probability that the lifespan is $> t$; more explicitly,

$$\bar{P}(t) = \int_t^\infty p(u)du. \qquad (1)$$

The mean lifespan μ is then given by the well-known formulae

$$\mu = \int_0^\infty tp(t)dt = \int_0^\infty \bar{P}(t)dt. \qquad (2)$$

In principle, if we could isolate a sample of newly produced RBCs released into the circulation (i.e., a cohort

sample), label them, and transfuse them into the recipient, the survival curve obtained by tracking the labeled RBCs would provide an estimate of the full SF $\bar{P}(t)$.

RBC compartmental model

We start with the LIDR compartmental model previously studied by several other groups [26, 36–38, 40]. In brief the model specifies that

$$\dot{r}(t) = k(t) - \int_{-\infty}^t k(u)p(t-u)du, \qquad (3)$$

where $r(t)$ denotes the number of cells in the RBC compartment at time t, $k(t)$ is the RBC production rate, which is the rate of entry of cells into the RBC compartment, and $p(t)$ is the full RBC lifespan PDF, which can be a completely arbitrary PDF on the positive real numbers. The integral term in Eq. (3) is the rate at which cells leave the compartment, i.e., the rate of RBC death due to senescence or hemolysis [40]. We assume $k(t)$ is bounded and non-negative, $0 \leq k(t) \leq M < \infty$ for all t, $-\infty < t < \infty$, and that the mean lifespan μ (Eq. (2)) is $< \infty$. Integrating Eq. (3) from 0 to $t > 0$, we obtain

$$r(t) - r(0) = \int_0^t k(u)du - \int_0^t k(v) \int_{-\infty}^v p(v-u)dudv.$$

Because $k(t)$ is bounded, both integrals are finite. After exchanging the order of integration in the double integral, letting $r_0 = r(0)$ be the size of the *index population* (i.e., the RBC population that was present at time $t_0 = 0$, which is the index time as described above), and some algebraic manipulation, we find

$$r(t) = r_0 + r_0(t) + r_1(t) - \int_0^\infty k(-u)\bar{P}(u)du, \qquad (4)$$

where

$$r_0(t) = \int_0^\infty k(-u)\bar{P}(t+u)du \qquad (5)$$

and

$$r_1(t) = \int_0^t k(u)\bar{P}(t-u)du. \qquad (6)$$

The integral in (4) represents the RBC population due to past production, that is, during $t < 0$, that is present at time $t_0 = 0$. This can be seen by noting that $k(-u)du$ is the number of cells produced during the time interval $-u$ to $-u + du$, and $\bar{P}(u)$ is the fraction of cells that survive for time $u > 0$, and the integral is the sum of these quantities over all $u > 0$. For a person with a stable internal environment, this will be the RBC population at time 0, thus r_0 and the integral cancel each other in Eq. (4) and we are left with

$$r(t) = r_0(t) + r_1(t). \tag{7}$$

The term $r_0(t)$ represents the part of the index population that has survived until time $t > 0$, whereas $r_1(t)$ is the portion of the RBC population that is present at time $t > 0$ that was produced between times 0 and t.

Residual lifespan distribution

The survival time after $t = 0$ of a cell in the index population is called the *residual lifespan* of the cell. The residual survival function (RSF), $\bar{P}_r(t)$, is the proportion of the index population that has survived at least until time $t > 0$ (equivalently, the probability of residual lifespan $> t$), $\bar{P}_r(t) = r_0(t)/r_0(0)$. From Eq. (5),

$$\bar{P}_r(t) = \frac{\int_0^\infty k(-u)\bar{P}(t+u)du}{\int_0^\infty k(-u)\bar{P}(u)du}. \tag{8}$$

Survival curves reported in studies of random RBC samples are in fact estimates of the RSF.

The RSF depends on the past production rate $k(t)$ for times $t < 0$, as indicated in Eq. (8). If $k(t)$ can be assumed constant for $t < 0$ (which is consistent with the assumption that the subjects had a stable intravascular environment), $\bar{P}_r$ reduces to the familiar form

$$\bar{P}_r(t) = \frac{1}{\mu}\int_t^\infty \bar{P}(u)du. \tag{9}$$

Note the similarity between Eqs. (9) and (1). The corresponding PDF of the residual lifespan is

$$p_r(t) = \bar{P}(t)/\mu. \tag{10}$$

Integration by parts of $\int_0^\infty \bar{P}_r(t)dt$ gives the *mean residual lifespan* μ_r, which is different from *mean full lifespan*:

$$\mu_r = \frac{E(\tau^2)}{2\mu} = \frac{\sigma^2 + \mu^2}{2\mu}, \tag{11}$$

where $E(\cdot)$ signifies expected value, and μ and σ^2 are the mean and variance of the full RBC lifespan τ.

Current age distribution

The members of the index population that have age $> t \geq 0$ at time 0 are those that were produced at some time u before time $-t$ and have survived for an additional time $-u$; as in the case of Eq. (4), this is given by

$$\int_{-\infty}^{-t} k(u)\bar{P}(-u)du = \int_t^\infty k(-u)\bar{P}(u)du.$$

The fraction of the index population of age $> t$, namely,

$$\bar{P}_c(t) = \frac{\int_t^\infty k(-u)\bar{P}(u)du}{\int_0^\infty k(-u)\bar{P}(u)du}, \tag{12}$$

is the SF of the *current age distribution*, which is the distribution of ages in the cell population at the index time $t_0 = 0$. It reduces to Eq. (9) if the past production rate is constant; in this specific situation, but not in general, the current age distribution is the same as the residual lifespan distribution and hence mean current age, $\mu_c =$ mean residual lifespan, μ_r.

Excess lifespan distribution

Consider the subpopulation of the index population that is still present at a specified time $t_e > 0$; these are the cells that were produced at some time $-u < 0$ and survived for a time $t_e + u$. The *excess lifespan distribution* corresponds to the SF $\bar{P}_e(t)$, which is the fraction of this subpopulation of age $> t_e + t$, where $t > 0$ is called the *excess lifespan*. By an argument similar to that following Eq. (4), the subpopulation of age $> t_e + t$ consists of those cells that were produced at some time some time $-u < 0$ and have survived for time $t_e + t + u$; the number of such cells is $\int_0^\infty k(-u)\bar{P}(t_e + t + u)du$. The entire subpopulation present at time t_e corresponds to $t = 0$ in this expression. Thus the fraction of those present at time t_e that are of age $> t_e + t$ is the ratio

$$\bar{P}_e(t) = \frac{\int_0^\infty k(-u)\bar{P}(t_e + t + u)du}{\int_0^\infty k(-u)\bar{P}(t_e + u)du}. \tag{13}$$

This is the conditional probability that the lifespan of an RBC in the index population is $> t + t_e$ given that it is $> t_e$. Given a constant past production rate it reduces to

$$\bar{P}_e(t) = \frac{\int_{t+t_e}^\infty \bar{P}(u)du}{\int_{t_e}^\infty \bar{P}(u)du} = \frac{\bar{P}_r(t_e + t)}{\bar{P}_r(t_e)}, \tag{14}$$

and this can be estimated from the observed residual survival curve. The corresponding PDF of excess lifespan distribution is given by

$$\bar{p}_e(t) = \frac{\bar{P}(t_e + t)}{\int_{t_e}^\infty \bar{P}(u)du}. \tag{15}$$

Note that, for $t_e = 0$, the excess lifespan SF is the same as the current age SF and consequently is the same as the residual lifespan SF in the case under consideration.

Relation (2) holds for any probability distribution, including the full lifespan, residual lifespan, current age, and excess lifespan distributions. We can use Eqs. (8), (12), and (14) to compute the mean residual lifespan (μ_r), mean current age (μ_c) in the index population, and the mean excess lifespan (μ_e) in the surviving part of the index population at any time $t_e > 0$. These results generalize those of Lindsell et al. [13], without the questionable assumption of a stable, stationary, and closed population [57].

Methods and statistical analysis

We used a nonlinear mixed effects (NLME) model [58] to analyze data obtained from 8 healthy adults who received autologous biotin-labeled RBCs [14]. Parametric bootstrap analysis was performed to confirm the results of NLME analysis because the number of subjects is small. The NLME model analysis was conducted using the *nlmefit* function in the Statistics toolbox of MATLAB software [59]. All computations and simulations were conducted using the same software. A personal computer with a 4 GB RAM and a 2 core Intel© (R) Core(TM) i3- M 370 @ 2.40 GHz processor was used for all computations.

We analyzed the lowest density bioRBC data from Mock et al. [14] as biotinylation at lower densities has negligible effect on the RBC lifespan [14]. Details of the experimental design and data collection are provided in [14]. Briefly, each of eight healthy subjects was transfused with autologous bioRBCs labeled at a biotinylation reagent density of 6 µg/mL of packed RBC. Post-transfusion residual RBC survival data were collected. Since even microliter blood samples contain many RBCs, the laws of large numbers from probability theory imply that such samples will replicate the residual lifespan distribution of the whole circulation at a given time within measurement error; this is the basis of the statistical analysis of random blood samples. Since the flow cytometric instrumentation allows tracking of the disappearance of at least 97 % of bioRBCs [14], we excluded values of the survival curves smaller than 0.03. The bioRBC concentration at day 1 was used as the baseline value.

We performed separate analyses for the Weibull, gamma, and lognormal models. For each model, the non-linear mean-value terms are the RSFs as in Eq. (9), but with the appropriate survival functions; thus,

$$\bar{W}_r(t;\alpha,\beta) = \frac{1}{\mu_w}\int_t^\infty \bar{W}(u;\alpha,\beta)du,$$
$$\bar{G}_r(t;\alpha,\beta) = \frac{1}{\mu_g}\int_t^\infty \bar{G}(u;\alpha,\beta)du, \qquad (16)$$
$$\bar{L}_r(t;\alpha,\beta) = \frac{1}{\mu_l}\int_t^\infty \bar{L}(u;\alpha,\beta)du,$$

where $\bar{W}_r$, $\bar{G}_r$, and $\bar{L}_r$ denote the RSFs, $\bar{W}$, $\bar{G}$, and $\bar{L}$ denote the full SFs, and μ_w, μ_g, and μ_l denote the mean full lifespans (see Appendix 1) in the Weibull, gamma, and lognormal models, respectively. The same letters α and β are used for all of the Weibull, gamma, and lognormal parameters for brevity in presentation, but we emphasize that they are entirely unrelated. We refer to α and β as the structural parameters on the original scale.

Just as $\bar{P}_r(t)$ is the ratio of the surviving part of the index population relative to the size of the index population, the observed residual survival curve, is the ratio of the measured concentration of the surviving part of the labeled RBCs at time $t > 0$ relative to the concentration at index time $t = 0$. In standard mixed effects model notation, the models in Eq. (16) are

$$\bar{W}_{ij} = \bar{W}_r(t_{ij};\alpha_i,\beta_i) + \epsilon_{ij},$$
$$\bar{G}_{ij} = \bar{G}_r(t_{ij};\alpha_i,\beta_i) + \epsilon_{ij}, \qquad (17)$$
$$\bar{L}_{ij} = \bar{L}_r(t_{ij};\alpha_i,\beta_i) + \epsilon_{ij},$$
$$i = 1,\ldots,M, \; j = 1,\ldots,n_i,$$

where $\bar{W}_{ij}$, $\bar{G}_{ij}$, and $\bar{L}_{ij}$ denote the observed residual survival curve corresponding to the Weibull, gamma, and lognormal models, respectively, $M = 8$ is the number of subjects, n_i is the number of observations on the ith subject, t_{ij} is the time of the jth observation on the ith subject, and ϵ_{ij} is an additive random error term, assumed normal $N(0,\phi^2)$ (we use ϕ^2 to denote variance of the random error to not confuse with σ^2 used to denote the variance of the full lifespan distribution). Thus $\bar{W}_{ij} = \bar{W}(t_{ij})$, for example, is the observed residual survival curve for the ith subject at the jth time using the Weibull model. To ensure the parameter estimates are > 0 we write

$$\alpha_i = \exp(a_i), \; \beta_i = \exp(b_i) \qquad (18)$$

and rewrite the models as

$$\bar{W}_{ij} = \bar{W}_r(t_{ij};a_i,b_i) + \epsilon_{ij},$$
$$\bar{G}_{ij} = \bar{G}_r(t_{ij};a_i,b_i) + \epsilon_{ij}, \qquad (19)$$
$$\bar{L}_{ij} = \bar{L}_r(t_{ij};a_i,b_i) + \epsilon_{ij},$$
$$i = 1,\ldots,M, \; j = 1,\ldots,n_i, \text{ and}$$

$$a_i = a_f + a_{ri}, \; b_i = b_f + b_{ri}. \qquad (20)$$

We call a_i and b_i the individual structural parameters on the log scale; this well known "log trick" [58] allows the use of unconstrained optimization techniques. $a_f \; (= \log(\alpha_f))$ and $b_f \; (= \log(\beta_f))$ denote the fixed effects, representing the average values of the parameters (in log scale) over the whole population from which the subjects in the sample are drawn (the reference population), and a_{ri} and b_{ri} denote random effects, which represent the deviations of an individual's parameters from the population average [58]. "Mixed effects" refers to the presence of both fixed and random effects in (20). The random effects vector (a_{ri}, b_{ri}) is modeled as multivariate normal, $N(\mathbf{0}, \Psi)$, independent of the measurement errors ϵ_{ij}, and independent between subjects; likewise, the errors are assumed independent between and within subjects [58].

For the distribution specified by the corresponding PDF, the derived parameters for each individual are the functions of corresponding α_i and β_i and for the population (fixed effects estimates) are the functions of corresponding α_f and

β_f. For the corresponding full lifespan distribution, the mean full lifespan (or simply mean lifespan) μ and the standard deviation σ are the derived parameters that can be computed easily using the formulas provided in Appendix 1. Other derived parameters describing RBC survival of potential clinical importance are

1. τ_{95}, defined as the 95th percentile of the full lifespan distribution, a surrogate for the maximum lifespan,
2. the mean current age, μ_c, defined as the mean of the current age distribution, which is equal to the mean of the residual lifespan distribution or the mean of the excess lifespan distribution at index time $t = 0$ assuming constant past production rate, and
3. the residual half-life, T_{50}, defined as the median of the residual lifespan distribution. Note that we use the term half-life irrespective of whether or not the distribution is exponential.

The values of τ_{95} and T_{50} have to be obtained numerically as they cannot generally be expressed in closed form; any modern statistical software has a facility to do this. An exception is that τ_{95} is easily obtained for the Weibull distribution: $\tau_{95} = \beta(\log(20))^{1/\alpha}$ (all logarithms in this paper are natural logarithms). The closed form expressions for μ_c for the corresponding PDFs are presented in Appendix 1. Formulas for mean residual lifespan and *steady state mean full lifespan* are provided in Appendix 1. Mathematical manipulations for efficient computation of the different survival functions based on the Weibull, gamma, and lognormal PDFs are also provided therein.

Results

Figure 1 shows the population fits for the 8 study subjects using the three different models. Residual survival data for each subject along with the respective individual fits using
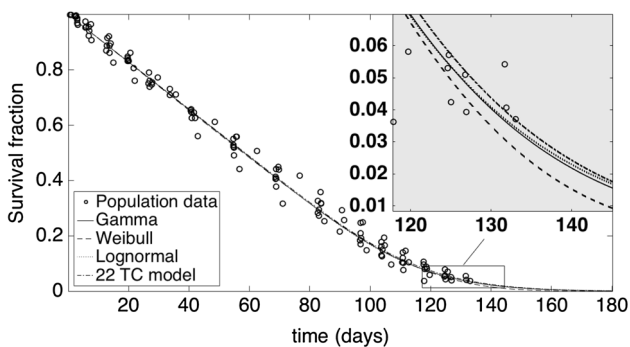

Fig. 1 Fitted population-level residual survival curves for data of all 8 study subjects. Inset: tail region where the models differ the most

the Weibull, gamma, and lognormal models are shown in Fig. 2. The group values of root mean square errors (RMSEs) are 0.0120, 0.0126, and 0.0129 for Weibull, gamma, and lognormal models, respectively. Corresponding fractional RMSE (RMSE/average response value) are 0.023, 0.024, and 0.025 respectively. The log likelihood (logl) values are 358, 354, and 351 respectively. The RMSE and logl values are shown in Table 1.

The individual weighted residuals, population weighted residuals, and conditional weighted residuals for all of the models were consistent with normality based on normal probability plots (not shown) and Shapiro–Wilk tests ($P \geq 0.05$). The population fit for the RSF along with the computed full lifespan, current age, and excess lifespan (at different times t_e after index time) SFs and corresponding PDFs using the gamma model are shown in Fig. 3 top and bottom panels respectively. Such a graph can easily be computed for individual subjects too (not shown). Similarly, the population and individual survival functions and the corresponding PDFs can easily be computed using the Weibull and lognormal models as well (not shown).

Population-level (fixed effects) estimates of the structural parameters on the original scale, α_f and β_f, for all the models are given in Table 2. For each of the three models, Table 3 gives maximum likelihood estimates (top), mean of the bootstrap estimates (middle), and the bootstrap 95 % CI of the population derived parameters μ, σ, τ_{95}, μ_c, and T_{50}. The predicted values of the individual structural parameters on original scale are shown in Table 4 and the values of the individual derived parameters of the RBC survival are shown in Table 5.

The range of individual μ was 97.5–128.0 days for the Weibull, 99.6–128.4 days for the gamma, and 100.95–128.37 days for the lognormal model. The range of σ was 18.84–36.97 days for the Weibull, 18.86–35.59 days for gamma, and 19.51–34.98 for the lognormal model. For τ_{95} the ranges were 149.2–160.1, 155.3–164.5, and 157.14–165.72 days, for μ_c they were 55.7–65.39, 56.18–65.59, and 48.62–63.50 days, and for T_{50} they were 50.1–64.0, 50.3–64.2, and 50.65–64.19 days, for the Weibull, gamma, and lognormal, respectively.

Discussion

Various distributions have been used for the full RBC lifespan in the literature, including the homogeneous lifespan model [2, 23, 25, 40], in which each of the cells has the same fixed lifespan, and the Weibull [22, 39], gamma [40], and lognormal [60] distributions. A more complicated model, a mixture of two Weibulls (a so-called

Fig. 2 Fitted individual-level residual survival curves for data of all 8 study subjects
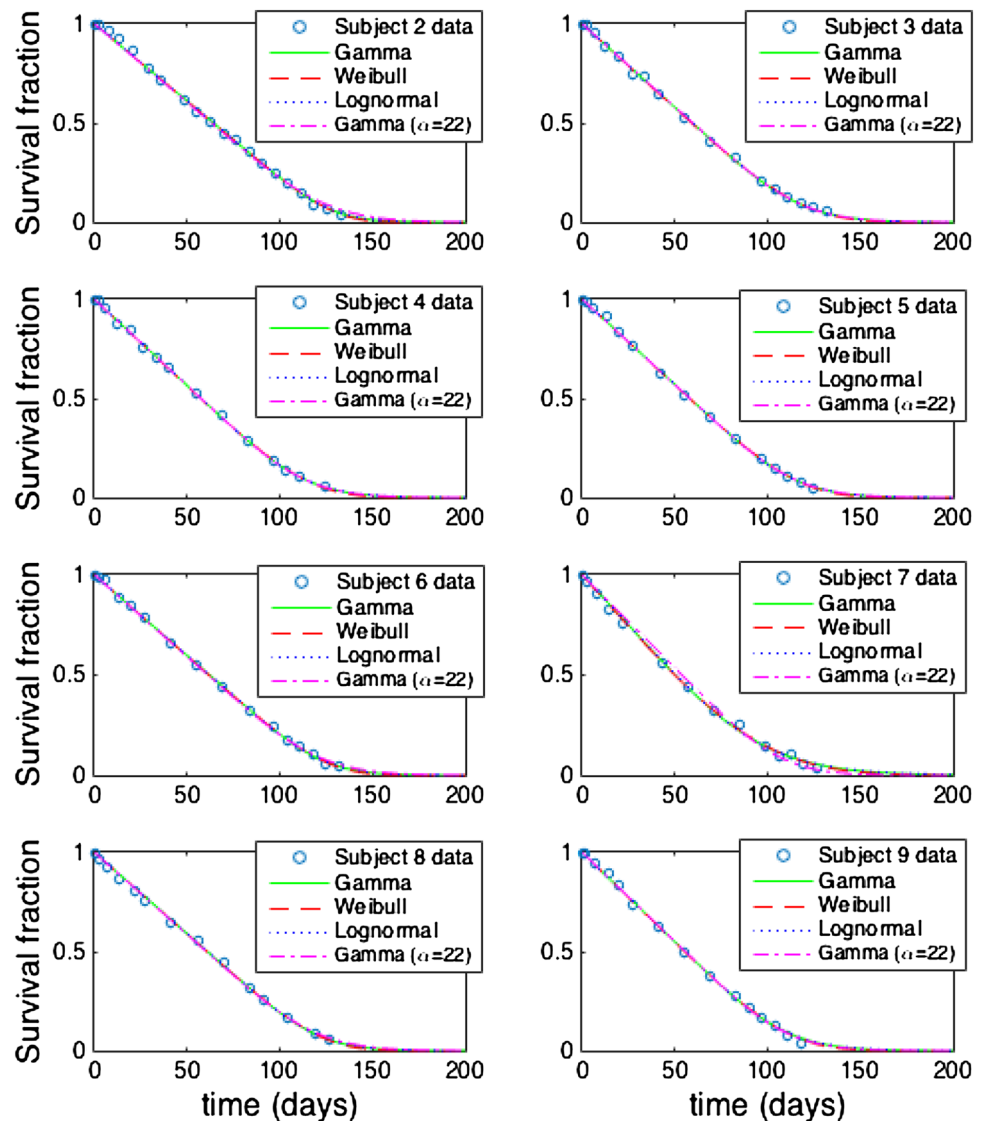
bathtub-shaped distribution), was used by Korell et al. [11, 19, 20, 30]. (Note that there is a misplaced minus sign in Eq. (1) in both [20] and [30]). By judicious choice of the parameters its shape can be made to mimic senescence, random destruction, and neocytolysis. This is a phenomenological model in the sense mentioned earlier.

In the present paper three separate models, using the Weibull, gamma, and lognormal distributions, were employed to analyze the bioRBC data from eight healthy adult subjects. The nonlinear mixed effects approach provided a framework for analysis of both population and individual variability of the structural and derived parameters. Because of the large number of parameters in their model, the authors in [11, 19, 20, 30] were unable to estimate all of them simultaneously and thus found it necessary to fix

several of them in the analysis. By contrast, we were able to carry out complete analyses with the Weibull, gamma, and lognormal models, each with only two structural parameters, without fixing any parameter in advance.

The fits appear excellent visually (Figs. 1 and 2). The fractional RMSEs (RMSE/ average response value) of 0.023, 0.024, and 0.025 respectively for Weibull, gamma, and lognormal models together with the fact that the residuals are consistent with normality support the use of the NLME methodology.

In addition to routine model diagnostics we performed a parametric bootstrap simulation (details in [21]) to confirm the results from the NLME software (MATLAB [59]), which are based on asymptotic theory for maximum likelihood and require the number of subjects to be large. The

Table 1 Goodness-of-fit and model selection criteria

Goodness of fit	Model		
	Weibull	Gamma	Lognormal
RMSE	0.0120	0.0126	0.0129
logl	358	354	351

RMSE root mean squared error, *logl* loglikelihood

bootstrap simulation utilizes the NLME information from the reference population to generate a large number of virtual datasets that duplicate the random mechanism that generated the original data set. The confidence intervals (CIs) of structural parameters obtained from the bootstrap analysis are similar to those from the NLME analysis. This shows that parameter estimation by the NLME software is reliable even with the small number of subjects ($M = 8$) in our dataset. Since analytical expressions for the standard errors (SEs) of the derived parameters are not available, we also used the bootstrap results to quantify their SEs in order to compute the resulting 95 % CIs (=estimate $\pm 1.96 \times$ SE).

All three models give physiologically plausible values for the parameters μ, σ, τ_{95}, μ_c, and T_{50} for both the population and the individual subjects. The agreement between

Fig. 3 Population survival functions (*top panel*) and PDFs (*bottom*) based on gamma model. For constant past production rate the residual lifespan, current age, and excess lifespan (for $t_e = 0$) SFs and PDFs coincide. Note the changes in excess lifespan SF and PDF depending on the choice of t_e

the three models is almost exact for μ, σ, μ_c, and T_{50}, but the gamma and lognormal models give consistently greater values (approximately 3 %) for τ_{95} than the Weibull. The reason is that the full lifespan distributions are very similar in their midranges but diverge towards the long survival tail, where the data are sparse, and τ_{95} is a property of the tail. We speculate that a data set with more reliable measurements in the tail region of the residual survival curve might provide a stronger basis for deciding which (if any) of the models is significantly better than the others. Techniques for more accurately enumerating bioRBCs after removal of 95 % of the initial population are currently being developed by some of the authors (JAW, DMM, PV-P).

Because of the small sample size there is no statistical criterion by which to decide confidently which is the "best" of the three models entertained here. For some purposes the choice of any particular model is immaterial, and in this case the three models are about equally mathematically and computationally tractable; see Appendix 1. One advantage of the gamma model in certain applications is that a clean state space representation exists when α is an integer. We presented such a gamma-based erythropoiesis model that is physiologically relevant and demonstrated its applicability using clinical data [61].

Recently, Lledo-Garcia et al. [9] gave extensive comparisons of three common RBC lifespan models: homogeneous lifespan, random destruction (RD), and transit compartment (TC) models. The TC model contains a series of compartments connected by first order cell transfer rates [27]. The TC model with one compartment is the RD model. A TC model with infinitely many compartments converges to the homogeneous lifespan model [27]. The TC model with 12 compartments was found by Lledo-Garcia et al. to describe the data best based on likelihood functions (in the form of the objective function value or OFV, defined as minus twice the loglikelihood) and visual predictive checks. The mean full lifespan for the 12 TC model was found to be 91.8 days for healthy subjects, which is substantially shorter than the normally accepted value of around 120 days.

There can be difficulties with likelihood-based model selection [62], especially for non-nested models [63] as occur in [9] and in the present paper. Unless the data satisfy the technical requirements, such as normality or independence, for several models, as must be determined from model diagnostics, a comparison based on likelihood may be questionable, as the likelihood function is then just a criterion function that is not the same for different models. Even when such requirements are met, it is not always clear what such a comparison means. To say, as is often done, that the model with the higher likelihood "explains
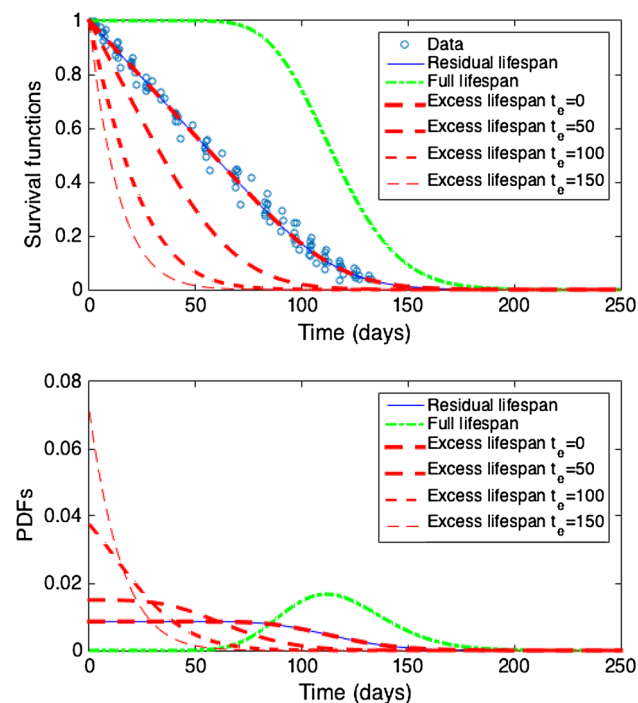
Table 2 Estimated population-level structural parameters α_f and β_f on original scale (subscript f indicates fixed effects)

Method	Parameter	Model		
		Weibull	Gamma	Lognormal
NLME	α_f	5.38	23.02	4.74
	95 % CI	(4.17, 6.58)	(13.99, 32.05)	(4.68, 4.80)
	Unit	–	–	–
	β_f	125.38	5.06	0.0421
	95 % CI	(119.75, 131.01)	(3.31, 6.81)	(0.0272, 0.0571)
	Unit	Day	Day	–
Bootstrap	α_f	5.42	23.54	4.74
	95 % CI	(4.32, 6.81)	(15.45,34.95)	(4.68, 4.80)
	Unit	–	–	–
	β_f	125.39	5.15	0.0428
	95 % CI	(119.88, 130.64)	(3.49,7.21)	(0.0282, 0.0593)
	Unit	Day	Day	–

The parameters in the Weibull, gamma, and lognormal models have different meanings and are neither related nor comparable. (i) Maximum likelihood estimates of the population fixed effects parameters on log scale are obtained from the NLME analysis; exponentials of the estimates give the values in original scale. Confidence intervals (CIs) on original scale are given by exp(fixed effect estimate on log scale $\pm$1.96SE) (ii) Bootstrap estimates of α_f and β_f are the means of the NLME estimates of 1000 virtual experiments

Table 3 Population estimates and confidence intervals for derived parameters

Derived Parameters	Parameter Info	Model		
		Weibull	Gamma	Lognormal
μ_f	NLME	115.60	116.57	116.64
	Bootstrap	115.62	116.71	116.79
	(CI)	(109.17, 121.66)	(110.81, 122.51)	(111.23, 122.75)
σ_f	NLME	24.77	24.30	24.19
	Bootstrap	25.84	24.38	24.28
	(CI)	(20.82, 28.81)	(20.53, 28.33)	(20.43, 27.73)
τ_{95f}	NLME	153.76	159.20	160.07
	Bootstrap	153.95	159.51	160.40
	(CI)	(150.02, 158.36)	(155.09, 164.00)	(156.00, 165.58)
μ_{cf}	NLME	60.45	60.82	57.26
	Bootstrap	60.51	60.93	57.32
	(CI)	(58.18, 62.85)	(58.77, 63.33)	(54.33, 60.61)
T_{50f}	NLME	57.94	58.29	58.32
	Bootstrap	57.97	58.36	58.40
	(CI)	(54.96, 60.90)	(55.45, 61.26)	(55.62, 61,37)

μ_f mean of the full lifespan distribution (or mean full lifespan), σ_f standard deviation of the full lifespan distribution, τ_{95f} 95th percentile of the full lifespan distribution, μ_{cf} mean of the current age distribution (or mean current age), and T_{50f} median of the residual lifespan distribution (or residual half-life). All the units are in days. Top: maximum likelihood estimate (computed from NLME estimates of structural parameters); middle: bootstrap estimate; bottom: 95 % bootstrap confidence interval. Subscript f indicates fixed effect

the data better" is circular, because, without some discussion of the mechanism behind the data, the only sense in which the data are "explained" in this situation is by the higher likelihood. Unfortunately the probabilistic justification of a comparison by likelihood will also be inapplicable if the sample size is small, as is the case in some of the studies discussed in [9] and here.

The gamma model in the current paper is equivalent to TC model with α compartments if α is a positive integer. For comparison we conducted NLME analyses of our bioRBC

Table 4 Estimated individual-level (mixed effects) structural parameters α_i and β_i on original scale

Model	Parameter	Subject								Mean	CV
		2	3	4	5	6	7	8	9		(%)
Weibull	α_i	8.07	4.99	5.33	5.38	6.62	2.86	6.41	4.99	5.58	27.24
	β_i	135.87	128.11	123.96	125.17	132.51	109.37	130.34	119.70	125.63	6.61
Gamma	α_i	46.38	21.01	21.99	22.79	33.92	7.83	31.51	19.28	25.59	45.13
	β_i	2.77	5.65	5.23	5.10	3.67	12.72	3.87	5.75	5.59	54.83
Lognormal	α_i	4.84	4.76	4.72	4.74	4.81	4.56	4.79	4.68	4.74	1.87
	β_i	0.023	0.042	0.045	0.042	0.029	0.113	0.032	0.052	0.047	59.42

Units are the same as in Table 2. The parameters in the Weibull, gamma, and lognormal models have different meanings and are neither related nor comparable. Subscript i indicates individual parameter

Table 5 Individual-level (mixed effects) estimates of derived RBC survival parameters

Model	Derived Parameter	Subject								Mean (days)	CV (%)
		2	3	4	5	6	7	8	9		
Weibull	μ_i	128.01	117.61	114.23	115.42	123.58	97.47	121.36	109.89	115.95	8.10
	σ_i	18.84	26.99	24.68	24.72	21.87	36.97	22.12	25.21	25.18	21.41
	τ_{95i}	155.65	159.64	152.30	153.48	156.40	160.51	154.67	149.15	155.22	1.41
	μ_{ci}	65.39	61.91	59.78	60.35	63.73	55.7	62.69	57.84	60.93	5.18
	T_{50i}	64.02	59.01	57.26	57.85	61.84	50.08	60.74	55.14	58.24	7.41
Gamma	μ_i	128.42	118.69	114.98	116.23	124.33	99.63	122.02	110.77	116.88	7.63
	σ_i	18.86	25.90	24.52	24.35	21.37	35.59	21.72	25.25	24.70	20.22
	τ_{95i}	160.93	164.24	158.05	158.96	161.40	164.48	159.83	155.27	160.39	1.93
	μ_{ci}	65.59	62.17	60.10	60.66	64.0	56.18	62.94	58.26	61.24	5.03
	T_{50i}	64.21	59.36	57.50	58.12	62.17	50.29	61.01	55.41	58.51	7.34
Lognormal	μ_i	128.37	119.30	115.14	116.46	124.47	100.95	122.09	110.97	117.22	7.31
	σ_i	19.51	24.69	24.61	24.20	21.36	34.98	22.04	25.55	24.62	18.81
	τ_{95i}	162.87	163.78	159.38	160.01	162.43	165.72	161.12	157.14	161.56	1.68
	μ_{ci}	63.50	58.56	56.47	57.16	61.41	48.62	60.17	54.29	57.52	8.04
	T_{50i}	64.19	59.66	57.58	58.23	62.24	50.65	61.05	55.49	58.64	7.36

μ_i mean full lifespan, σ_i standard deviation of the full lifespan, τ_{95i} 95th percentile of the full lifespan distribution, μ_{ci} mean current age, T_{50i} residual half-life. All the units are in days. Subscript i indicates individual parameter

data with 1–50 compartments, namely, by taking $\alpha = 1$ to 50 and estimating only β in each case. The smallest RMSE (= 0.0151), fractional RMSE (= 0.029), and largest log-likelihood (= 336), all occurred for the 22-compartment TC model ($\alpha = 22$), as shown in Fig. 3. These values were less favorable than those for the unconstrained Weibull, gamma, and lognormal models; in particular, the loglikelihood for the lognormal model (= 351), which was the smallest log-likelihood for our three models, was considerably larger than that of the 22-compartment TC model.

It is accepted in the literature that RBC lifespan values are tightly dispersed around a mean value for healthy individuals. There are many diseases which impact RBC

lifespan, e.g., sickle cell disease, diabetes, chronic kidney disease, etc. In such cases, information about the whole distribution of RBC lifespan, in particular the mean (μ) and the standard deviation (σ), may be helpful in distinguishing between health and disease. A simultaneous estimation of μ and σ of the RBC lifespan distribution has never been done to the best of our knowledge. Typically, RBC survival studies have been focused on quantifying an average description, e.g., mean age (i.e mean current age according to our definition) of circulating RBCs [3], RBC survival (meaning the half life of the ^{51}Cr disappearance from the circulation) [5], half-life [4], mean lifespan [9, 11, 20, 64], mean potential lifespan (MPL) [14], mean remaining life

span [12] (this is similar to mean excess lifespan at $t_e = 0$ according to our definition), etc.

The concepts of residual lifespan, current age, and excess lifespan have often appeared in the literature, but are not named explicitly and sometimes not treated rigorously [3, 8, 33, 53, 54, 57]. Outside of RBC survival literature, the concepts of current age and excess lifespan appear in renewal theory [65] but have not previously been used in compartmental models, which are conceptually different from renewal theory.

In the literature, the maximum RBC lifespan T_{max} is defined as the time when all of the labeled RBCs disappear from the circulation [3, 11, 13]. Taken literally, this means T_{max} is the time at which the residual survival curve decreases to zero. The determination of that time is highly dependent on the sensitivity of the method and associated analytical instrumentation because the proportion of labeled RBCs becomes very small towards the tail region. T_{max} is often estimated as the time at which a linear or nonlinear curve fitted to the residual survival curve intersects the time axis. When a linear fit of the entire residual survival curve is used, this intersection is actually an estimate of the mean lifespan rather than the maximum [31, 33]. When a nonlinear fit is used [3, 13], the estimate is strongly dependent on the last sample time of the residual survival curve. The final 2 points of the residual survival curve were extrapolated to the time axis to estimate the maximum RBC lifespan in [3]; such an estimate will also depend on the distance between the last two measurement points.

We used the 95th percentile τ_{95} of the lifespan distribution as a surrogate for the maximum lifespan; by definition, τ_{95} is the value such that 95 % of the RBC lifespans are $< \tau_{95}$. (The Weibull, gamma, and lognormal distributions extend to infinity in the positive direction; hence have no true maximum value). The choice of 95 % is arbitrary and is used here only for illustration. Any other percentile could be used; in practice the choice would likely be dictated by the nature of the application. The estimated population values of τ_{95} are in physiologically plausible ranges.

Recently, a new clinically relevant parameter to assess the quality of transfused red blood cells called mean remaining lifespan (MRL) has been introduced [12]. MRL is analogous to the area under the curve (AUC) or mean residence time (MRT) in PK studies. It is defined as the AUC of the fraction of the transfused RBCs remaining in the circulation versus time [12]. This is in fact an estimate of mean residual lifespan described in the current paper ($\mu_r = \int_0^\infty \bar{P}_r(t)dt = \mu_c$). Measurement of MRL as described in the paper [12] again runs into the problem of determining when the last of the transfused RBCs have left the circulation. For practical purposes, MRL is replaced by

$MRL_{0.95}$, which is defined as the area under the curve until the time, $t_{0.95}$, when 95 % of the transfused cells have disappeared from the circulation. $t_{0.95}$ is estimated by interpolation of the curve [12]. The accuracy of $MRL_{0.95}$ depends on the frequency of sampling around $t_{0.95}$ and the last sample must wait until at least 95 % of the transfused cells have disappeared, which takes approximately 3–4 months in healthy individuals.

As shown by Dornhorst [33], it follows from Eq. (9) that the mean RBC lifespan μ is given by the negative reciprocal of the slope of the RSF at $t = 0$, a result that holds for any lifespan distribution, assuming the past production rate is constant. Thus the MPL (mean potential lifespan) described in [14] is actually an estimate of μ for each individual (note that we use the data from [14]). The average MPL for the eight subjects given in [14] was 115 ± 8 days, which is very close to our population values of $\mu = 116 \pm 3$ for Weibull and 117 ± 3 for both the gamma and lognormal (mean $\pm$ SE). Similarly, the time to disappearance of 50 % of the labeled RBCs from the circulation, denoted by T_{50} and given as 58 ± 4 days in [14], is very close to the population residual half-lives T_{50} found here (58 ± 1.5 for all models).

Why use a complicated method like NLME when the simple MPL method gives similar results? There are both statistical and practical reasons. From the statistical point of view, first, the average MPL in [14] is the average of the individual MPLs obtained by simple linear regression for each subject separately. This approach yields an overestimate of the variability among individuals, and hence the SD of the MPLs, based on only 8 subjects, is likely too large. By contrast, NLME gives estimates and standard errors at both the population and individual levels [58]. Second, the MPL does not model the full lifespan distribution and, therefore, is incapable of providing quantitative information about other survival parameters such as the longest lifespans in the distribution as reflected in τ_{95}. Information about the whole distribution may be helpful in distinguishing between health and disease (e.g., sickle cell disease). Third, the NLME framework provides standard methodology to include covariates such as sex, age, ethnicity, etc. in the analysis. Fourth, NLME is well-suited to the situation of sparse samples, which is usually the case in clinical settings. The results of a NLME analysis can be used to determine the minimum number, optimal timing, and the last time point of measurement for measurements in a new subject when only a few measurements are possible [21, 66], e.g., in infants or sick patients.

Our results in Appendix 1 allow complete, rigorous, and computationally efficient analyses of the models considered here (parameter estimation took under a second for Weibull and just over a second for the gamma and lognormal models

on a personal computer). Further, there are no constraints on parameters during the estimation process, as in other published models [9, 20, 30], or computational constraints, such as the inability of the software to handle more than 30 compartments described by Lledó-García et al. [9].

Conclusions

Starting with definitions of index population, which has previously not been made explicit, and of birth time, death time, full lifespan, residual lifespan, current age, and excess lifespan, we provided mathematical descriptions of RBC survival parameters, which remove the lack of clarity often found in the literature. We exhibit the connections between a compartmental (or LIDR) model for the RBC population based on a given lifespan distribution and the residual lifespan, current age, and excess lifespan distributions, not available previously in the literature. We gave analytical expressions for mean full lifespan, mean residual lifespan, mean current age, and steady-state mean full lifespan, and indicated how to compute RBC half-life and 95th percentile of the RBC lifespan distribution (τ_{95}). The use of τ_{95} (or other percentiles) avoids the questionable concept of T_{max} often used in the literature.

Employing nonlinear mixed effects modelling, we analyzed residual survival data from biotin-labeled RBCs using models based on the Weibull, gamma, and lognormal distributions. The three models fit the data closely and gave equally physiologically plausible estimates of clinically interpretable RBC survival parameters at population and individual levels.

Our modelling framework could be useful in studying RBC lifespans in various diseases that affect RBC survival, especially in situations with strongly non-linear survival curves (e.g., sickle cell anemia). The model cannot be used if the assumption of "stability" of the subject's internal environment is not satisfied, as may occur in the case of blood loss and/or other significant intercurrent events (e.g., events leading to hemolysis). The framework also lends itself to analyzing richer data sets containing covariates such as age, gender, and weight.

Appendix 1: Weibull, gamma, and lognormal PDFs

The PDF of the Weibull distribution, denoted by $w(t; \alpha, \beta)$, is given by

$$w(t; \alpha, \beta) = \frac{\alpha}{\beta} \left(\frac{t}{\beta} \right)^{\alpha-1} e^{-(t/\beta)^{\alpha}}. \tag{21}$$

The Weibull parameters α and β are both > 0 and are called the shape and scale parameters, respectively. The mean and variance for the Weibull are

$$\mu_w = \beta \Gamma \left(1 + \frac{1}{\alpha} \right), \ \sigma_w^2 = \beta^2 \left[\Gamma \left(1 + \frac{2}{\alpha} \right) - \left(\Gamma \left(1 + \frac{1}{\alpha} \right) \right)^2 \right]. \tag{22}$$

The PDF of the gamma distribution with parameters $\alpha, \beta > 0$ is

$$g(t; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} t^{\alpha-1} e^{-t/\beta}, \quad t > 0 \tag{23}$$

(Γ denotes the Euler gamma function). Note that we write g instead of p for the PDF of the gamma distribution. The parameters α and β are again called the shape and scale parameters, respectively. The mean and variance are

$$\mu_g = \alpha\beta, \ \sigma_g^2 = \alpha\beta^2. \tag{24}$$

Let $X \sim N(\alpha, \beta^2)$, i.e., normal with mean α and variance β^2, and $Y = e^X$. Then Y has the lognormal distribution with parameters α and β, and PDF

$$l(t; \alpha, \beta) = \frac{1}{t\beta\sqrt{2\pi}} \exp \left(-\frac{(\ln t - \alpha)^2}{2\beta^2} \right), \ t > 0. \tag{25}$$

The mean and variance for the lognormal are

$$\mu_l = \exp \left(\alpha + \frac{\beta^2}{2} \right), \sigma_l^2 = \exp(2\alpha + \beta^2)(\exp(\beta^2) - 1). \tag{26}$$

The same letters α and β are used for the Weibull, gamma, and lognormal parameters for brevity in presentation, but we emphasize that the parameters in the three distributions are entirely unrelated. The PDFs represent the full RBC lifespan distributions in the Weibull, gamma, and lognormal models respectively and the means correspond to the full RBC lifespan.

Appendix 2: Survival functions for Weibull, gamma, and lognormal lifespan distributions

For any lifespan distribution we have

$$\int_t^\infty \bar{P}(u)du = \int_t^\infty \int_u^\infty p(v)dv du. \tag{27}$$

Interchanging the order of integration yields

$$\int_t^\infty vp(v)dv - t\bar{P}(t). \tag{28}$$

The *full survival functions* (Eq. (1)), in the case of gamma, Weibull, and lognormal lifespan distributions are given by

$$\bar{G}(t;\alpha,\beta) = \int_t^\infty g(v;\alpha,\beta)dv = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_t^\infty v^{\alpha-1}e^{-v/\beta}dv, \tag{29}$$

$$\bar{W}(t;\alpha,\beta) = e^{-\left(\frac{t}{\beta}\right)^\alpha}, \tag{30}$$

and

$$\bar{L}(t;\alpha,\beta) = 1 - N\left(\frac{log(t)-\alpha}{\beta}\right) \tag{31}$$

respectively, where $N(x)$ is the CDF for the standard normal $Z \sim N(0,1)$. Thus using Eqs. (9) and (28), we obtain the *residual survival function* in the case of the gamma lifespan distribution

$$\bar{G}_r(t) = \bar{G}_c(t) = \bar{G}(t;\alpha+1,\beta) - \frac{t}{\alpha\beta}\bar{G}(t;\alpha,\beta). \tag{32}$$

Similarly, from Eq. (14) we obtain the *excess lifespan survival function* for the gamma lifespan distribution

$$\bar{G}_e(t) = \frac{\alpha\beta\bar{G}(t_e+t;\alpha+1,\beta) - (t_e+t)\bar{G}(t_e+t;\alpha,\beta)}{\alpha\beta\bar{G}(t_e;\alpha+1,\beta) - t_e\bar{G}(t_e;\alpha,\beta)} \tag{33}$$

Given α and β, the survival functions in Eqs. (32) and (33) are readily calculated by numerical integration in most statistical packages, including MATLAB [59]; Eqs. (32) and (33) show that the survival functions of the residual lifetime and the current and excess lifespan distributions are no more difficult to calculate than the full survival function. It also follows from Eqs. (32) and (2) that the mean residual lifespan and mean current age are both equal to $(\alpha+1)\beta/2$: the integral from 0 to ∞ of $\bar{G}(t;\alpha+1,\beta)$ equals $(\alpha+1)\beta$ and the integral of $t\bar{G}(t;\alpha,\beta)$ equals $\alpha\beta^2(1+\alpha)/2$, as shown using integration by parts, $u = \bar{G}(t;\alpha,\beta)$, $dv = tdt$.

The survival functions $\bar{G}_r$ and $\bar{G}_c$ of the residual lifetime and current age distributions, for the gamma lifespan distribution cannot be expressed in closed form, but there is a closed form when α is an integer ≥ 1. The gamma distribution with $\alpha = 2$ has been used [61] in an erythropoiesis model for chronic kidney disease patients.

Although the Weibull survival function (30) has a simple form, the corresponding residual survival function does not. The integral in Eq. (9) can be expressed in terms of the gamma survival function:

$$\int_t^\infty e^{-\left(\frac{u}{\beta}\right)^\alpha}du = \frac{\beta}{\alpha}\int_{\left(\frac{t}{\beta}\right)^\alpha}^\infty v^{\left(\frac{1}{\alpha}-1\right)}e^{-v}dv$$
$$= \beta\Gamma\left(1+\frac{1}{\alpha}\right)\bar{G}\left(t^\alpha;\frac{1}{\alpha},\beta^\alpha\right). \tag{34}$$

The residual, current age, and excess lifespan survival functions, $\bar{W}_r$, $\bar{W}_c$, and $\bar{W}_e$, can now be written for the Weibull model as follows:

$$\bar{W}_r(t;\alpha,\beta) = \bar{W}_c(t;\alpha,\beta) = \bar{G}(t^\alpha;1/\alpha,\beta^\alpha), \text{ and} \tag{35}$$

$$\bar{W}_e(t;\alpha,\beta) = \frac{\bar{G}((t_e+t)^\alpha;1/\alpha,\beta^\alpha)}{\bar{G}(t_e^\alpha;1/\alpha,\beta^\alpha)}. \tag{36}$$

Using Eqs. (9), (14), (28) and (31), the residual, current age, and excess lifespan survival functions, $\bar{L}_r$, $\bar{L}_c$, and $\bar{L}_e$ for the lognormal model can be written as follows:

$$\bar{L}_r(t;\alpha,\beta) = \bar{L}_c(t;\alpha,\beta)$$
$$= 1 - N\left(\frac{log(t)-\alpha}{\beta}-\beta\right)$$
$$- \frac{t}{\mu_l}\bar{L}(t;\alpha,\beta), \text{ and} \tag{37}$$

$$\bar{L}_e(t;\alpha,\beta) = \frac{1 - N\left(\frac{log(t_e+t)-\alpha}{\beta}-\beta\right) - \frac{(t_e+t)}{\mu_l}\bar{L}(t_e+t;\alpha,\beta)}{1 - N\left(\frac{log(t_e)-\alpha}{\beta}-\beta\right) - \frac{t_e}{\mu_l}\bar{L}(t_e;\alpha,\beta)}. \tag{38}$$

From Eqs. (2) and (9) the mean residual lifespans for the gamma, Weibull, and lognormal distributions are given by $\mu_{g,r} = \frac{(\alpha+1)\beta}{2}$, $\mu_{w,r} = \frac{\beta\Gamma(2/\alpha)}{\Gamma(1/\alpha)}$, and $\mu_{l,r} = \frac{\exp(\alpha+3\beta^2/2)}{2}$, respectively (using Eq. (11)); these are obviously different from the corresponding mean full lifespans (Appendix 1). We also note that, in equilibrium, the steady state residual lifespan distribution has PDF $p_{r,ss}(t) = \bar{P}(t)/\mu$, which is the same as the transient residual lifespan distribution. The steady state full lifespan has PDF $p_{ss}(t) = tp(t)/\mu$ [60]. Thus the mean residual and mean full lifespans in steady state are respectively $(\sigma^2+\mu^2)/2\mu$ (Eq. (11)) and $(\sigma^2+\mu^2)/\mu$. In steady state, therefore, the mean residual lifespan is one half the mean full lifespan. For the gamma, Weibull, and lognormal the steady state mean full lifespans equal $\mu_{g,ss} = (\alpha+1)\beta$, $\mu_{w,ss} = 2\beta\Gamma(2/\alpha)/\Gamma(1/\alpha)$, and $\mu_{w,ss} = \exp(\alpha+3\beta^2/2)$, respectively. For an individual in stable condition, these distributions are what would be seen for RBCs in the circulation.

References

1. Ashby W (1921) The determination of the length of life of transfused blood corpuscles in man. J Exp Med 29:267–281
2. Krzyzanski W, Perez-Ruixo JJ, Vermeulen A (2008) Basic pharmacodynamic models for agents that alter the lifespan

distribution of natural cells. J Pharmacokinet Pharmacodyn 35:349–377

3. Cohen RM, Franco RS, Khera PK, Smith EP, Lindsell CJ, Ciraolo PJ, Palascak MB, Joiner CH (2008) Red cell life span heterogeneity in hematologically normal people is sufficient to alter HbA1c. Blood 112(10):4284–4291

4. Ly J, Marticorena R, Donnelly S (2004) Red blood cell survival in chronic renal failure. Am J Kidney Dis 44(4):715–719

5. Vos FE, Schollum JB, Coulter CV, Doyle TC, Duffull SB, Walker RJ (2011) Red blood cell survival in long-term dialysis patients. Am J Kidney Dis 58:591–598

6. Banks HT, Bliss KM, Tran H (2012) Modeling red blood cell and iron dynamics in patients with chronic kidney disease. Int J Pure Appl Math 75

7. Franco RS (2009) The measurement and importance of red cell survival. Am J Hematol 84:109–114

8. Luten M, Roerdinkholder-Stoelwinder B, Schapp NPM, de Grip WJ, Bos HJ, Bosman GJCGM (2008) urvival of red blood cells after transfusion:a comparison between red cells concentrates of different storage periods. Transfusion 48(7):1478–1485

9. Lledó-García R, Kalicki RM, Uehlinger DE, Karlsson MO (2012) Modeling of red blood cell life-spans in hematologically normal populations. J Pharmacokinet Pharmacodyn 39(5):453–462

10. Callender ST, Powell EO, Witts LJ (1945) The lifespan of the red cell in man. J Pathol Bacteriol 57(1):129–139

11. International Committee for Standardization in Hematology (1980) Recommended method for radioisotopic red-cell survival studies. Br J Haematol 45:659–666

12. Kuruvilla DJ, Nalbant D, Widness JA, Veng-Pedersen P (2014) Mean remaining life span: a new clinically relevant parameter to assess the quality of transfused red blood cells. Transfusion 41(10pt2):2724–2729

13. Lindsell CJ, Franco RS, Smith EP, Joiner CH, Cohen RM (2008) A method for the continuous calculation of the age of labeled red blood cells. Am J Hematol 83:454–457

14. Mock DM, Matthews NI, Zhu S, Strauss RG, Schmidt RL, Nalbant D, Cress GA, Widness JA (2011) Red blood cell (RBC) survival determined in humans using RBCs labeled at multiple biotin densities. Transfusion 51:1047–1057

15. Brown IW Jr, Eadie GS (1953) An analytical study of in vivo survival of limited populations of animal red blood cells tagged with radioiron. J Gen Physiol 36(3):327–343

16. Eadie GS, Brown IW Jr (1953) Analytical review: red blood cell survival studies. Blood 8(12):1110–1136

17. Eadie GS, Brown IW Jr, Curtis WG (1955) The potential life span and ultimate survival of fresh red blood cells in normal healthy recipients as studied by simultaneous Cr^{51} tagging and differential hemolysis. J Clin Investig 34(4):629–636

18. Garby L, Mollison PL (1971) Deduction of mean red-cell lifespan from ^{51}Cr survival curves. Br J Haematol 20(1):527–536

19. Korell J, Coulter CV, Dufful SB (2011b) A statistical model for red blood cell survival. J Theor Biol 268:39–49

20. Korell J, Vos FE, Coulter CV, Schollum JB, Walker RJ, Dufful SB (2011c) Modeling red blood cell survival data. J Pharmacokinet Pharmacodyn 38:787–801

21. Shrestha RP (2012) Modeling the lifespan of red blood cells. PhD Dissertation, University of Massachusetts., Amherst, MA

22. Freise KJ, Schmidt RL, Widness JA, Veng-Pedersen P (2008a) Pharmacodynamic modeling of the effect of changes in the environment on cellular lifespan and cellular response. J Pharmacokinet Pharmacodyn 35:527–552

23. Ramakrishnan R, Cheung WK, Wacholtz MC, Minton N, Jusko WJ (2004) Pharmacokinetic and pharmacodynamic modeling of recombinant human erythropoietin after single and multiple doses in healthy volunteers. J Clin Pharmacol 44:991–1002

24. Sun YN, Jusko WJ (1998) Transit compartments versus gamma distribution function to model signal transduction processes in pharmacodynamics. J Pharm Sci 87:732–737

25. Uehlinger DE, Gotch FA, Sheiner LB (1992) A pharmacodynamic model of erythropoietin therapy for uremic anemia. Clin Pharmacol Ther 51(1):76–89

26. Krzyzanski W, Ruixo JJP (2012) Lifespan based indirect response models. J Pharmacokinet Pharmacodyn 39:109–123

27. Krzyzanski W (2011) Interpretation of transit compartments pharmacodynamic models as lifespan based indirect response models. J Pharmacokinet Pharmacodyn 38:179–204

28. Hamren B, Bjork E, Sunzel M, Karlsson M (2008) Models for plasma glucose, HbA1c, and hemoglobin interrelationships in patients with type 2 diabetes following tesaglitazar treatment. Clin Pharmacol Ther 84(2):228–235

29. Glynn SA (2001) The red blood cell storage lesion: a method to the madness. Transfusion 50:1164–9

30. Korell J, Coulter CV, Duffull SB (2011a) Evaluation of red blood cell labelling methods based on a statistical model for red blood cell survival. J Theor Biol 291:88–98

31. Mock DM, Lankford GL, Widness LF, Burmiester DK, Strauss RG (1999) Measurement of red cell survival using biotin-labeled red cells: validation against ^{51}Cr-labeled red cells. Transfusion 39:156–162

32. Brown GM, Hayward OC, Powell EO, Witts LJ (1944) The destruction of transfused erythrocytes in anaemia. J Pathol Bacteriol 56(1):81–94

33. Dornhorst AC (1951) The interpretation of red cell survival curves. Blood 6:1284–1292

34. Schiødt E (1938) On the duration of life of the red blood corpuscles. Acta Med Scand 95(1):49–79

35. Krzyzanski W, Ramakrishnan R, Jusko WJ (1999) Basic pharmacodynamic models for agents that alter production of natural cells. J Pharmacokinet Biopharm 27(5):467–489

36. Al-Hunity NH, Widness JA, Schmidt RL, Veng-Pedersen P (2005) Pharmacodynamic analysis of changes in reticulocyte subtype distribution in phlebotomy-induced stress erythropoiesis. J Pharmacokinet Pharmacodyn 32(3):359–376

37. Pérez-Ruixo JJ, Krzyzanski W, Hing J (2008) Pharmacodynamic analysis of recombinant human erythropoietin effect on reticulocyte production rate and age distribution in healthy subjects. Clin Pharmacokinet 47(6):399–415

38. Freise KJ, Widness JA, Schmidt RL, Veng-Pederson P (2007) Pharmacodynamic analysis of time-variant cellular disposition: reticulocyte disposition changes in phlebotomized sheep. J Pharmacokinet Pharmacodyn 34:519–547

39. Freise KJ, Widness JA, Schmidt RL, Veng-Pederson P (2008b) Modeling time variant distributions of cellular lifespans: increases in circulating reticulocyte lifespans following double phlebotomies in sheep. J Pharmacokinet Pharmacodyn 35:285–323

40. Krzyzanski W, Woo S, Jusko WJ (2006) Pharmacodynamic models for agents that alter production of natural cells with various distributions of lifespan. J Pharmacokinet Pharmacodyn 33(2):125–166

41. Barlow RE, Proschan F (1975) Statistical theory of reliability and life testing: probability models. Holt, Rinehart and Winston, New York

42. Meeker WQ, Escobar LA (1998) Statistical methods for reliability data. Wiley, New York

43. Nachlas JA (2005) Reliability engineering: probabilistic models and maintenance methods. Taylor and Francis, Boca Raton

44. Lee ET (1980) Statistical methods for survival data analysis. Wadsworth, Belmont

45. Birnbaum ZW, Saunders SC (1958) A statistical model for life-length of materials. J Am Stat Assoc 53(281):151–160

46. Castorina P, Blanchard P (2007) Unified approach to growth and aging in biological, technical and biotechnical systems. SpringerPlus 1:7

47. Gavrilov LA, Gavrilova NS (1991) The biology of life span: a quantitative approach. Harwood Academic Publisher, New York

48. Gavrilov LA, Gavrilova NS (2001) The reliability theory of aging and longevity. J Theor Biol 213:527–545

49. Juckett DA, Rosenberg B (1993) Comparison of the Gompertz and Weibull functions as descriptors for human mortality distributions and their intersections. Mech Ageing Dev 69:1–31

50. Antonelou MH, Kriebardis AG, Papassideri IS (2010) Aging and death signalling in mature red cells: from basic science to transfusion practice. Blood Transfus 8(3):s39–s47

51. Greenwalt TJ, Dumaswala UJ (1988) Effect of red cell age on vesiculation in vitro. Br J Hematol 68:465–467

52. Handelman GJ, Levin NW (2010) Red cell survival: relevance and mechanism involved. J Ren Nutr 20(5):S84–S88

53. Tinmouth A, Fergusson D, Ian Yee CP, Hébert C, Investigators A, the Canadian Critical Care Trials Group (2006) Clinical consequences of red cell storage in the critically ill. Transfusion 46(11):2014–2017

54. Franco RS (2012) Measurement of red cell lifespan and aging. Transfus Med Hemother 39:302–307

55. Gifford SC, Derganc J, Shevkoplyas SS, Yoshida T, Bitensky MW (2006) A detailed study of time-dependent changes in human red blood cells: from reticulocyte maturation to erythrocyte senescence. Br J Haematol 135:395–404

56. Franco RS (1998) Time-dependent changes in the density and hemoglobin F content of biotin-labeled sickle cells. J Clin Investig 101(12):2730–2740

57. Müller H, Wang J, Carey JR, Caswell-Chen EP, 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 3:125–131

58. Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer, New york

59. MATLAB (2010) MATLAB and Statistics Toolbox Release 2010a. Mathworks Inc., Natick

60. Dowling MR, Milutinović D, Hodgkin PD (2005) Modeling cell lifespan and proliferation: is likelihood to die or to divide independent of age? J R Soc Interface 2:517–526

61. Chait Y, Horowitz J, Nichols B, Shrestha RP, Hollot CV, Germain MJ (2014) Control-relevant erythropoiesis modeling in end-stage renal disease. IEEE Trans Biomed Eng 61(3):658–64

62. Cavanaugh JE (2015) Model selection lecture XIV: the application of model selection criteria. url: http://myweb.uiowa.edu/cavaaugh/ms_lec_14_ho.pdf. Accessed Jun 2015

63. Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57:307–333

64. Mitlyng BL, Singh JA, Furne JK, Ruddy J, Levitt MD (2006) Use of breath carbon monoxide measurements to assess erythrocyte survival in subjects with chronic diseases. Am J Hematol 81:432–438

65. Parzen E (1967) Stochastic processes. Holden-Day, San Francisco

66. Merrill S, Horowitz J, Traino AC, Chipkin SR, Hollot CV, Chait Y (2011) Accuracy and optimal timing of activity measurement in estimating absorbed dose of radioiodine in the treatment of graves' disease. Phys Med Biol 56:1–11