**Biomarkers of healthy ageing: expectations and validation**

Carmen Martin-Ruiz and Thomas von Zglinicki*

Institute for Ageing and Health, Newcastle University, Campus for Ageing and Vitality, Newcastle upon Tyne NE4 5PL, UK

The challenge of devising a set of biomarkers capable of measuring the ageing rate in human subjects was articulated long ago. In recent years, progress in the basic biology of ageing suggests the realistic possibility of preventive or restorative interventions that may extend healthy lifespan in mammals including human subjects. Specifically, frailty is being increasingly recognised as a clinically relevant syndrome that may be therapeutically addressed. This greatly enhances the need for sensitive and specific biomarkers of healthy ageing that are validated in both experimental animals and, importantly, in human subjects over the whole age range. Here, we will discuss the present challenges and requirements for biomarker validation in human subjects. We propose the central requirements for a validated biomarker of healthy ageing as: (i) better predictive power than chronological age for multiple dimensions of ageing; (ii) identification of the age range in which the marker is informative; (iii) establishment of sensitivity/specificity as indicators of its predictive power at the level of the individual; (iv) minimisation of methodological variation between laboratories.

### Biomarkers and the stochastic nature of the ageing process

Over the last 30 years, biogerontology has moved from an observational to an interventional science with increasingly realistic potential for human interventions. This has generated an urgent need for markers that can precisely predict the biological age of populations, groups and individuals. Various approaches to define criteria for biomarkers of ageing, either in conjunction with or opposed to biomarkers of age-related disease have been published\(^1\)–\(^4\). The essential feature of a biomarker of ageing was defined by Baker and Sprott\(^2\) as ‘a biological parameter that either alone or in some multivariate composite will’ … ‘better predict functional capability at some late age, than will chronological age’, although the impact of age-related disease as originally excluded by Baker and Sprott is still a matter of debate\(^1\)–\(^7\). Extensive programmes to validate marker candidates for intervention testing in mice\(^7\) and non-human primates\(^5\) have been run, however, with limited success so far\(^6\). This is to a large extent due to our still insufficient mechanistic understanding of the ageing process. Ageing is immensely complex. It is to a significant extent governed by chance, leading to stochastic distributions of all parameters that define the rate of ageing even in genetically identical individuals under (as much as possible) identical environmental influences\(^8\). While we know many of the gene products and environmental influences and their principal routes of interactions that determine the rate of ageing, the impact of any of these on the ageing process in a given individual can vary greatly due to chance events that may occur already during early development. In some cases this will be ‘true’ chance that is by its nature unpredictable (as described by the uncertainty principle in particle physics). In other cases, it will be randomness that arises from the sheer number of interactions each of which is essentially deterministic (and so can, in principle, be measured and assessed). Finally, experimental ignorance, not having

*Corresponding author: T. von Zglinicki, email t.vonzglinicki@ncl.ac.uk

**Abbreviations:** SBP, systolic blood pressure; TL, telomere length.

\(^1\)Baker and Sprott (2) defined a biomarker of ageing as:

\(\ldots\) a biological parameter that either alone or in some multivariate composite will better predict functional capability at some late age, than will chronological age.\)

\(^2\)Baker and Sprott (2): 'a biological parameter that either alone or in some multivariate composite will better predict functional capability at some late age, than will chronological age'.
discovered the relevance of a gene or the existence of a pathway, is still a major cause of unexplained variance in ageing. For all these reasons, we are yet far from understanding ageing mechanistically. This is reflected by the fact that there is no definition of ageing as a process that occurs in an individual. Rather, the best available definition of biological ageing is a probabilistic one, by which ageing is identified by an ever increasing intrinsic probability of death with progressing time.

Accordingly, the perfect biomarker would allow the precise measurement of the probability of death at any given time. It is immediately clear from the above that a truly perfect biomarker of ageing cannot exist because there will always be true or apparent chance events in the future that change the ageing trajectory of an individual. In other words, biomarkers of ageing are by their nature probabilistic with a limited precision of a prediction at the level of the individual. Conversely, every improvement of biomarker prediction contributes to the reduction of apparent randomness and ignorance. Therefore, biomarkers of ageing not only have a utilitarian value, but do contribute greatly to the conceptual understanding of the ageing process.

**Ageing and disease or what is a biomarker of healthy ageing?**

There has been a longstanding debate in the field whether biomarkers of ageing should (and could) specifically measure basic underlying processes independent of age-related disease\(^1\)\(^{-}\)\(^7\). It has often been proposed that ageing, being a basic process underlying the development of disease and frailty, should be researched (and biomarkers of ageing should be validated) in disease-free subjects (for review see\(^9\)). A biomarker of healthy ageing in this sense would indicate underlying ageing biology not modified by disease. This concept was highly relevant in the early days of biogerontology when the field struggled to prove that ageing was more than and distinct from the sum of age-related morbidities. However, it is now well established that common basic biological processes, typically triggered by molecular damage and modified by cellular and systemic responses drive ageing at the level of the organism and modify risks for multiple diseases in a tissue-, organ- and system-specific manner. In turn, disease will feed back into the underlying molecular networks and thus impact onto the rate of ageing as well as enhance the risk for additional disease. For instance, chronic inflammation, which is strongly associated with most age-related chronic diseases, including dementia, depression, atherosclerosis, cancers, and diabetes, aggravates cellular senescence, which in turn reduces tissue regenerative potential and enhances pro-inflammatory signals, potentially increasing the risk for additional diseases\(^10\)\(^{-}\)\(^1\)\(^1\). There is also good epidemiological evidence that prevalence of a chronic disease is a significant risk factor for incidence of additional, often multiple age-associated degenerative diseases (see for instance\(^15\)). Thus, both from an opportunistic and a conceptual point of view it seems appropriate to view ‘basic’ ageing and age-related multimorbidity and disability as a continuum, especially with respect to biomarker development and validation.

Therefore, a biomarker of healthy ageing should not be confounded with a parameter that is informative or discriminatory for the healthy part of the ageing population only. Rather, it is a parameter that predicts the probability for maintenance of health with increasing age with high specificity and sensitivity.

Maintenance of health in an ageing population is a relative concept; at ages 85 years and above there is essentially no one free severe disease, and multimorbidity is common\(^11\). It is also a multidimensional concept, including not only absence of disease/multimorbidity but also cognition, capability/dependency, frailty and, ultimately, longevity. In the old, correlations between most of these dimensions (especially those to multimorbidity) are weak, suggesting that individual biomarkers will have different predictive power for multiple dimensions of the ageing process. For instance, there has been a longstanding debate as to whether and to what extent lifespan and healthspan co-vary in the old\(^14\)\(^{-}\)\(^16\). Multimorbidity, disability and mortality are at best loosely associated in octa- and nonagenarians\(^17\).

However, all these dimensions are associated with and driven by the ageing process. We therefore proposed\(^17\) that an informative biomarker of (healthy) ageing should be predictive for several of these dimensions. To better capture the multidimensionality of the ageing process, additional endpoint measures, prominently including measures of psychological and mental well-being, need to be considered for biomarker validation, as these deteriorate in significant subgroups of the population with important consequences for physiology and perception of the ageing process.

The concept of frailty requires special consideration in the context of ageing biomarker validation. Frailty is characterised by increased vulnerability to stress resulting in an increased risk of adverse health outcomes including disability, hospitalisation, institutionalisation and death. There is as yet no universally accepted definition of frailty. The two leading concepts are frailty as a clinical syndrome; a cluster of specific symptoms and signs including weight loss, exhaustion, low physical activity, muscle weakness and slow walking speed as developed by Fried et al.\(^18\) or as a cumulative index of health deficits and indicator of biological age as proposed by Rockwood et al.\(^19\)\(^{-}\)\(^20\). These individual deficits can include diseases, symptoms, signs, function tests and laboratory tests. Provided enough deficits are included in the index, their exact nature seems unimportant\(^21\).

Thus, frailty can be regarded as a complex biomarker of ageing (the Rockwood model) or as a clinical definition of an ageing syndrome (the Fried model), clearly illustrating the ambivalence between endpoints and biomarkers in ageing research. Application of both models to the same population shows that they measure overlapping but not identical concepts with significant fractions of participants falling in one but not the other frailty category. Interestingly, on a cohort level, associations
to a large number of biomarker candidates, especially inflammation markers, were very similar for both frailty models\(^{23}\).

**Use of birth cohorts for biomarker validation**

Chronological age is the most universally available ‘biomarker’ of ageing (forensics being a notable exception). However, it is also a weak marker; the differences in survival between the longest- and the shortest-living member of a cohort are typically greater than mean or median lifespan of the cohort, even in genetically and environmentally homogeneous cohorts under protected conditions with very little impact of external causes of death. Therefore, the first requirement for a candidate biomarker of ageing is that it needs to have better predictive power than chronologic age. Many population-based studies include participants over a wide age range, and associations are adjusted for age. This might not always be a robust procedure, given that biomarker candidates and their predictive power are often non-linearly associated with chronologic age (see later). This problem is circumvented if associations between marker candidates and ageing ‘outcomes’ are analysed in birth cohorts, in which all cohort members fall within a narrow age range. This approach addresses directly the relevant question, namely: are (groups of) individuals that are of the same chronologic age different in their ‘biological age’ and if so, by how much?

The limitation of the birth cohort approach is that, even if the study group was representative for the whole population at that age, it answers the question only for a narrow age group. There is now ample evidence that the predictive power of multiple candidate biomarkers of ageing varies with age group. For instance, up to about age 70 years systolic blood pressure (SBP) increases continuously with age\(^{23}\) and high SBP is a well-recognised risk factor for CVD and associated mortality\(^{24}-26\). At higher ages, however, SBP decreases rather than increases with age\(^{23}\) and higher blood pressure becomes protective in terms of all-cause mortality and cognition, whereas low SBP confers increased risks for mortality, cognitive impairment and disability\(^{17,27-31}\). Similarly, short peripheral blood telomere length (TL) is recognised as a risk factor for both mortality and (multi-) morbidity\(^{32-37}\). These associations are strongest in the age group up to about 75 years but tended to disappear in older populations\(^{34,38}\). Similar decreases of predictive power at higher age have been noted for other potential biomarkers\(^{39,40}\), although often there are not sufficient data on multiple age cohorts, especially the oldest old, available. Cohort and/or period effects may be partially responsible if there is a trend reversal or loss of predictive power at old age. Typically, later born cohorts are physically and cognitively healthier and show extended life expectancy as compared to earlier born cohorts at the same chronological age\(^{50}\). Increasingly widespread use of certain medication in the older population will influence biomarker associations. For instance, we found an association between high levels of vitamin D and cognitive impairment in a population-based study of 85-year olds, which is most probably explained by vitamin D supplementation specifically in care homes\(^{17}\). An increased use of anti-hypertensive medication in this age group may partially explain the decrease of SBP. However, the general pattern remained even for participants not on anti-hypertensive medication and after adjustment for survivor bias\(^{23}\). Moreover, low SBP predicted increased mortality in 90-year olds without heart failure, defined by low levels of N-terminal prohormone of brain natriuretic peptide\(^{30,41}\).

Life history cohorts are the ideal test bed to establish the age dependency of candidate biomarkers of ageing. Life history cohorts are birth cohorts for which candidate marker information is available longitudinally over a large fraction of the complete life history and which have reached a sufficient age to be informative about age-related outcomes. There are at least 60–70 human ageing cohorts that have been studied longitudinally worldwide\(^{42,43}\); however, very few of these qualify as life history cohorts. Examples of the latter from the UK include the MRC National Survey of Health and Development\(^{44}\) and the Lothian Birth Cohorts of 1921 and 1936\(^{45,46}\); see also www.halcyon.ac.uk. Some candidate biomarkers of physical capability (grip strength), cardiovascular function (SBP) and cognition have in fact been longitudinally assessed for long periods of time, with follow-ups spanning in some cases over 50 years in the same participants, enabling comprehensive validation of their predictive power over the life course\(^{23,47-56}\). However, for the vast majority of biomarker candidates, life-course longitudinal data are not available and will not be for a long while, if at all.

For instance, TL as a biomarker of ageing was only introduced in 2000\(^{33}\). In this case, the best validation strategy follows the biomarker criteria derived by Nakamura et al.\(^{3,5}\) by combining longitudinal analyses in multiple birth cohorts, in which the longitudinal change with age is expected to be consistent with the cross-sectional differences between the cohorts. However, in human subjects life history is strongly dependent on year of birth, and life expectation increases with time\(^{57}\). In addition, if biomarker data from multiple cohorts are pooled, technical variation between different laboratories becomes a concern. So far, longitudinal biomarker studies have seldom if ever been done in multiple cohorts performed by a single laboratory with no variation in methodology. For instance, we measured peripheral blood cell TL in about 7000 participants of six UK cohorts with consistent methodology (C Martin-Ruiz, T von Zglinicki and HALCYON Study team, unpublished results). However, technical variation in blood sampling and DNA extraction could not be avoided, and the observed cohort-specific differences could thus not definitely be attributed to variation of average biological age between cohorts. It should be noted that Nakamura’s criteria\(^{45}\) also request the rates of age-related change of a potential biomarker to be proportional to differences in the ageing rate or lifespan among the related species. It might be concluded that most of ageing marker candidates in present use have not been sufficiently validated.
How to validate a candidate biomarker of healthy ageing?

As discussed earlier, the main problem arising from the complex and stochastic nature of the ageing process is that there is no good single process or parameter to test biomarker prediction against. In other words, there is no gold standard. In most animal and a large number of human ageing biomarker studies, survival or lifespan is used as the closest approximation to an estimate of ageing rates. While this is in keeping with the definition of biological ageing, it has two major disadvantages: it rapidly loses power in small cohorts, and as discussed earlier, its association to ‘health span’ is uncertain. We believe that four steps are necessary for validation of a candidate biomarker of healthy ageing: (i) Prove better predictive power than chronological age for multiple dimensions of ageing; (ii) Identify the informative age range; (iii) Establish sensitivity/specificity; (iv) Assess (and minimise) methodological variation between laboratories.

We discussed earlier the essentiality to cover the multidimensionality of the ageing process, the advantages of birth cohorts for biomarker candidate validation and the problems associated with use of multiple birth cohorts to establish the informative age range. However, the two latter points deserve some further comments.

So far, the majority of studies in the ageing biomarker field do not go beyond establishing correlations at the cohort level. However, even a highly significant correlation between a biomarker and an outcome in a large cohort does not necessarily imply that the biomarker in question will predict the outcome with any degree of certainty for an individual. To give an example, we tested the ability of measuring TL in stroke survivors immediately after the stroke to predict incidence of dementia and 2-year survival\(^{(58)}\). In a linear model, every 1000bp of TL resulted in a decreased hazard ratio (HR) for incidence of dementia (HR=0.19, 95% CI 0.19-0.54, \(P=0.002\)). However, performing a receiver operator curve analysis, it was found that despite these strong associations a telomere setting that would result in the correct prediction for 80% of those that developed dementia would also predict 44% false positives. This resulted in a total of only 59% correct predictions (as compared to 50% by pure chance) because only a minority of patients did develop dementia\(^{(58)}\). Both the advance of preventive and restorative interventions into ageing processes (e.g. frailty) and the increasing commercialisation of biomarker services (e.g. telomeres) make it essential to generate this type of information as part of the validation process of biomarkers of ageing.

It is generally assumed that the relation between a biomarker and the outcome it stands for can be described by a linear function (possibly after some simple mathematical transformation of the marker values). Due to the complexity of the ageing process and the uncertain associations between its multiple domains discussed earlier, this will often not be true for candidate biomarkers of ageing. This is illustrated in Fig. 1, showing the distribution of TL in peripheral blood in a population-based birth cohort sample of 2660 participants aged 53 years. From a total of eleven cohorts comprising 9929 participants and spanning an age range from 50 to 88 years, we obtained a linear regression between age and TL with very narrow confidence intervals as

\[
TL = 8359 \text{ (sem 77)} - 58 \text{ (sem 1)} \times \text{Age}
\]

Assuming that the same equation would also describe the distribution of ‘biological age’ according to TL within the 53-year-old cohort, we would find a ‘biological age’ for the participants with a TL representing the median of reasonable 48 years. For the upper and lower quartiles of the distribution ‘biological ages’ of 28-6 and 70-2 years, respectively, would be calculated.

However, the upper and lower deciles of the telomere distribution, still representing 266 participants each, would end up with calculated ‘biological ages’ of less than 4 or more than 85-5 years, which is clearly unrealistic. Thus, the association between the biomarker TL and biological age is definitely different from the one between average TL and chronological age, most probably non-linear and possibly not even a continuous function (e.g. only a small part of the biomarker variation between individuals may be informative for any given domain of ageing).

Finally, the use of any potential biomarker is severely restricted as long as methodological variation between laboratories has not been independently assessed and...
minimised. Again using peripheral blood TL as an example for many candidate biomarkers of ageing, it is clear that at present no general applicable reference ranges for a ‘normal’ TL at any given age can be defined because data differ so widely between laboratories. This prohibits direct pooling of data from different laboratories to increase power as necessary for genotype-phenotype association studies. It also implies that measurements of this biomarker in individuals are useless as long as they cannot compared to a reference set established in the same laboratory.

Some recent biomarker examples

Recent progress in high-throughput, ‘-omics’ technologies has enabled unbiased searches for novel candidates for biomarkers of healthy ageing to start. However, despite the clear heritability of longevity, genome-wide association studies have so far largely failed to deliver novel marker candidates, probably both due to the very complex genomics involved in the ageing process and to the problems of defining the phenotype of healthy ageing clearly. A number of interesting associations with metabolomic and lipidomic parameters have been found, but the marker suffers from low specificity and sensitivity and the limited methodological variation between laboratories as discussed earlier. High levels of p16 (CDKN2A) are also an indicator of cell senescence. p16 expression was first suggested as a biomarker of ageing in mice in 2006 and first applied to human subjects in 2006. Its informative age range and whether it is actually better than chronological age still needs to be established. Persistent DNA damage may trigger either apoptosis or cell senescence, both of which may be associated with ageing. DNA damage also induces the phosphorylation of the histone variant H2AX (then called γH2AX), which forms foci at sites of DNA damage, especially double-strand breaks. γH2AX has recently been put forward as a potential candidate biomarker of ageing with clinical potential. The capacity to repair DNA double-strand breaks is impaired with age. γH2AX foci have been used for detection and clinical assessment of tumours in human subjects (for review see), they increase with age and they have been used as a marker for morbidity and age-related diseases; however, from an epidemiological point of view this marker candidate is yet far from being validated as most of the studies are small and the methodologies applied are not completely compatible. Cell senescence is now known to be mechanistically integrated with inflammation and with oxidative stress. However, so far a few markers of molecular oxidative damage or inflammation appear useful and consistent as biomarkers of ageing in human cohorts.

A candidate biomarker of ageing should only be regarded as fully validated if it fulfils the requirements stated earlier. These are very stringent criteria that are only just met by some of the longest established biomarkers of ageing, hand grip strength being a notable example. This test has strong potential as a screening tool because of its simplicity and its robust association with disability, mortality, and health care indicators such as longer hospitalisation or risk of post-surgery complications. In healthy adults, lower hand grip associates with all-cause mortality and increased risk of disability in later life over a wide age range. However, the association between mortality and hand grip strength becomes weaker in cohorts aged 60 years and over and, as evident in studies with long-term follow up, might be modified by cohort effects. Other tests of physical performance such as ‘Functional reach’, ‘Timed Up and Go’, and ‘One-leg stance’, also predict frailty, disability and mortality in the short term on those over 65 but appear to lose power in the very old. Even for such relatively simple assessments of physical performance, comparability between studies is limited by methodological concerns. In any case, the predictive power of these and other ‘classical’ biomarkers of ageing is low and their sensitivity/specificity does not reach the limits required for a biomarker with diagnostic power, fuelling the ongoing need for novel marker candidates.

Acknowledgements

We are grateful to all cohort members and study researchers involved in the specific studies referred in this paper.

Financial Support

The work was funded by the UK Medical Research Council [grant numbers G0601333 and G0500997] and the NIHR Newcastle Biomedical Research Centre in Ageing and Chronic Disease. The funders had no role in the design, analysis or writing of this article.

Conflicts of Interest

None.

Authorship

T. von Z. and C. M. M. designed and wrote the paper.
References


