

## Telomere Length and Telomere-related Genetic Variations in Epidemiology: Getting the Context Right

Fernando Pires Hartwig\*

Postgraduate Program in Epidemiology, Federal University of Pelotas; Oncology Research Group, Technology Development Center (Biotechnology Unit), Federal University of Pelotas, Brazil

### Abstract

Telomeres (5' TTAGGG 3' DNA *tandem* repeats and associated proteins) are in the spotlight of biomedical research given the body of evidence linking them with physiological aging and aging-related diseases, including tissue self-renewal failure (and consequences of this in several organs) and cancer. The epidemiological method has been extensively applied to study telomere biology in mainly two contexts: telomere length measurement and genotyping genetic variants associated with telomere length. Although both methods have the same common goal of better understanding the roles of telomeres in health and disease, they greatly differ regarding applications and limitations. In this manuscript, both methods are compared regarding common issues in causal inference: reverse causation, confounding, effect mediation and discerning a causal factor from a biomarker. In conclusion, telomere length measurement and genotyping genetic proxies can be combined to increase robustness, although some applications required the dynamics of telomere length, thus requiring robust study designs. In addition, the use of an appropriate conceptual framework can assist both data collection and analysis in many situations.

**Keywords:** Telomere length; Genetic proxies; Aging; Causality; Conceptual framework

**Abbreviation:** TL: Telomere Length

### Introduction

Since the discovery of telomerase in *Tetrahymena thermophila* [1,2], research on telomere biology has been receiving a great deal of attention, resulting in the Nobel Prize award for telomerase discovery [3]. Telomeres are 5' TTAGGG 3' DNA *tandem* repeats, with spatial structure organized by at least six associated proteins, called shelterin [4]. In linear chromosomes, telomere length (TL) shortens after each cell division due to the end-replication problem, resulting in telomere shortening with time when telomerase is lacking. In multicellular organisms, telomere shortening is evidenced as a fundamental aspect of aging. In this context, it has been shown that adult stem cells have telomere activity levels sufficient only to delay telomere shortening after a cell division, indicating that even these cells eventually reach a critical TL state. Since critical telomere shortening and/or loss of telomere capping activate apoptosis and senescence responses, TL is considered one of the main mechanisms of tissue renewal failure in old age [5,6], which is corroborated by the phenotypes seen in telomere syndromes [7]. Telomere biology also has a critical role in cancer: telomere maintenance is required for indefinite cell divisions, which is mainly achieved by telomerase activity (present in 85-90% of human cancers) [8], although a recombination-based (telomerase-independent) mechanism, termed alternative lengthening of telomeres, can also occur [9]. In addition, genomic instability (which is highly associated with telomere dysfunction) is considered a cancer hallmark as an early event in the carcinogenic process [10].

The relevance of telomeres in human diseases associated with aging (including cancer and impairments caused by loss of tissue renewal capacity) combined with the recent trends of the age structure of the majority of the countries worldwide makes telomere biology a very interesting study object for several biomedical research fields [11,12], including epidemiological research. Applying the epidemiological method to telomere biology can aid to the understanding of both the causal roles of the telomeres in human diseases or other health outcomes and how environmental exposures can affect TL, with

important implications for etiological research and development of telomere-related interventional strategies. In this regard, there are two main approaches for studying telomere biology at the population level: one is to actually measure TL; the other is to genotype genetic variants involved in TL regulation. Although the two approaches might seem to be two different but redundant means for studying telomere biology, they have quite different characteristics, including advantages, disadvantages and applications of each. Given the amount of attention that telomere biology has been receiving in scientific (including epidemiological) research, this manuscript intends to summarize these differences in order to assist researchers to determine the best study design for the purposes of the investigation in question as well as to make correct inferences and interpretations of the literature on this topic.

One of the major challenges of epidemiological research in general is how confidently a causal inference can be drawn from observed associations. In Figure 1, associations that are likely to be tested for significance are depicted. In this simplified scheme, interest lies in studying the relationship between TL and health outcomes ( $e_1$ ). To this end, TL measurement and/or genotyping telomere-related genetic variations can be performed. Based on this scheme, some key points in epidemiological research of telomere biology are discussed.

### Telomere Length x Genetic Proxies of Telomere Length

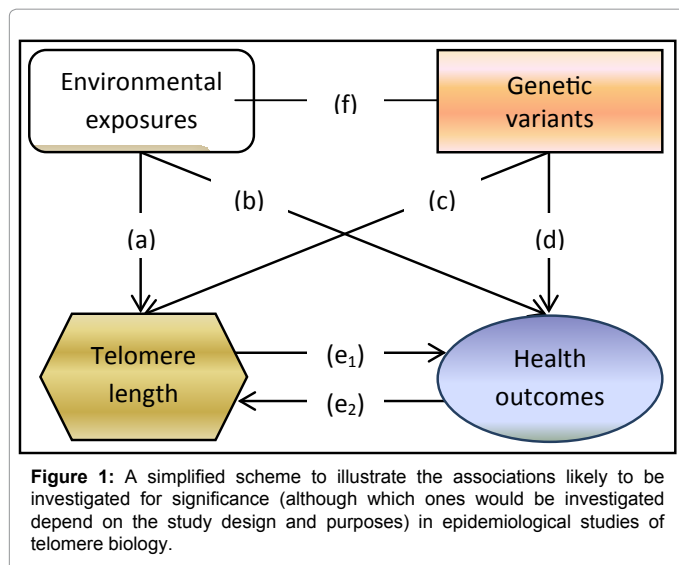
To study telomere biology epidemiologically, the most obvious method is to measure TL in different groups of individuals. Normally

\*Corresponding author: Fernando Pires Hartwig, Postgraduate Program in Epidemiology, Federal University of Pelotas; Oncology Research Group, Technology Development Center (Biotechnology Unit), Federal University of Pelotas, Brazil, Tel: (5553)81347172; E-mail: [fernandophartwig@gmail.com](mailto:fernandophartwig@gmail.com)

Received May 31, 2013; Accepted June 19, 2013; Published June 22, 2013

Citation: Hartwig FP (2013) Telomere Length and Telomere-related Genetic Variations in Epidemiology: Getting the Context Right. J Genet Syndr Gene Ther 4: 150. doi:10.4172/2157-7412.1000150

Copyright: © 2013 Hartwig FP. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



TL is measured in easily accessible tissues, such as peripheral blood cells or cells from the oral mucosa [13]. Obviously, the consequences of some exposures (that might be of interest regarding telomere biology) on TL in these cells will be different than in other cell compartments, indicating the need of knowing how representative TL of blood or buccal cells is of TL of other cell compartments. In this context, it is also important to ponder whether TL measurement in accessible tissues, for the outcome under study, is a factor directly involved in disease etiology or a biomarker of disease risk/status. Although TL is much more stable than, for example, serum glucose levels, it is still a biomarker that is subjected to some classical epidemiological difficulties for causal inference, such as reverse causation and confounding.

To overcome these difficulties of observational studies, an alternative that has been employed in different contexts in epidemiology is an approach called Mendelian randomization, which uses genetic variants as proxies of the exposure of interest. In the case of TL, obvious proxies would be genetic variants in genes that are directly linked to TL regulation (such as genes that encode telomerase subunits or other telomere-related proteins, some of which are causative of telomere syndromes as dyskeratosis congenita [7,14]), although new variants have been identified in genome-wide association studies [15-20], which are perhaps more likely to be useful as TL proxies given that some dyskeratosis congenita genetic profiles, for example, tend to be very rare. The main advantage of using TL genetic proxies instead of actually measuring TL lies on Mendel's 2<sup>nd</sup> Law of independent assortment, which implicates that the alleles of different loci randomly segregate during meiosis. What follows is that the genotype an individual inherits from his or her parents is independent, in principle, of any environmental aspects (Mendelian randomization was comprehensively reviewed, including its limitations, elsewhere [21,22]).

### Reverse Causation

Reverse causation is a frequent issue when information on both exposure and outcome are collected at the same time (e.g., in cross-sectional designs). In the context of telomere research, if the outcome of interest can influence TL ( $e_2$ ), causal inference from studies using this type of design is to be taken with caution (including the reasoning based on a theoretical framework [23]) or, more conservatively, an association found in this design (unknown if it is either  $e_1$  or  $e_2$ ) can

be further assessed for causality in more appropriate designs (e.g., in longitudinal studies). Two published studies with different designs illustrate this concept well. In a large retrospective case-control study in China, an association with telomere length and type 2 diabetes was found, even after controlling for potential confounders [24]. However, there is evidence that premature cell senescence and oxidative stress are both causes and consequences of type 2 diabetes [25,26], as well as for a role of oxidative stress in telomere shortening, which is implicated in cell senescence [27,28]. Such evidence indicates that the observed association between TL and type 2 diabetes is strongly subjected to reverse causation, which was recognized by the authors themselves. On the other hand, a recent longitudinal study of a 5-year nutritional intervention found significant associations between both baseline and dynamic TL and obesity [29]. In this case, reverse causation is not an issue for causal inference due to the strength of the longitudinal design.

In this context, using genetic variations associated with TL would provide a robust alternative regarding reverse causality, since germline genetic profiles always meet the temporality criteria. In case a genetic proxy is used and an association with the outcome (d) is observed, it can be inferred that this outcome is associated with TL (without temporality issues). Of note, if TL and outcome can influence each other (both  $e_1$  and  $e_2$  occur), it is expected to observe a different association between TL and exposure (a combination of  $e_1$  and  $e_2$ ) than what would be expected based on the associations between TL genetic proxies and the outcome (d), and between TL (c), which would consist of an estimate of  $e_1$  only.

### Confounding

Confounding occurs when there is a variable(s) that is (are) associated with both the exposure and the outcome, but is not in the causal pathway of the relationship exposure  $\rightarrow$  outcome ( $e_1$ ). This scenario results in this "third" variable interfering with the association between exposure and outcome. Considering that an environmental exposure is associated with both TL (a) and outcome (b) – and it is not the case that  $e_1$  is mediated by such exposure – such associations are likely to influence  $e_1$  (which can be established based on temporality discrimination provided by the study design and/or theoretical considerations). Although confounding is a potential issue in all studies that do not randomize the exposure, it is unlikely that all possible confounding variables are available for a given study. As an example, a review on the roles of TL in atherosclerosis cited some studies that provided evidence for a causal role of smoking (which is a known risk factor for atherosclerosis) in telomere shortening [30]. Assuming the authors have not considered other pieces of literature (especially regarding mediating roles of TL in the causal pathway of risk factors and atherosclerosis), these relationships would be expected to result in an association between TL and atherosclerosis in observational studies. However, it would be uncertain whether such association would be due to confounding or to an actual causal role of TL as a mediating factor for disease. Evidently, randomizing the exposure is expected to robustly protect TL from confounding; such design was recently applied to study the impacts of omega-3 polyunsaturated fatty acid supplementation on TL [31].

Given these considerations, replacing or combining TL measurement with genotyping TL genetic proxies would provide an alternative robustly protected from confounding, given Mendel's 2<sup>nd</sup> Law of independent assortment. Nevertheless, confounding is still a possibility (although much less likely) even when using genetic proxies, since associations between environmental exposures and genetic

variants (f) can occur either by mere chance and, more importantly, in the presence of population stratification. While the chance associations issue is likely to be solved by comparing different studies, population stratification has to be taken into account as a source of systematic error [21,22]. This indicates that for both telomere length measurement and TL genetic proxies genotyping there is possibility of confounding (although the probabilities greatly different between these methods), thus reinforcing the need of basing the data collection on an appropriate conceptual framework.

### Telomere Length as an Outcome/Mediating Factor

Using a conceptual framework also facilitates to discern confounding from effect mediation. Such discernment can be also done statistically. Comparing the results of modeling the association between TL and outcome ( $e_1$ ) and modeling an environmental exposure with TL (a) and outcome (b) with modeling  $e_1$  controlling for the same independent variables simultaneously can result in two main situations: loss of  $e_1$ , with both a and b still significant; or loss (or attenuation) of b, with  $e_1$  and a still significant. While the former indicates that the environmental exposure is a confounder of  $e_1$ , the later would indicate that TL mediates (at least partially) the effects of the exposure on the outcome (b), in a way that  $b = e_1 + TL\text{-independent effects}$ . However, it is important to note that the complexity of this reasoning increases according to the number of variables in the model, indicating that a conceptual framework is of great value in discerning confounding and effect mediation. In addition, identifying a conceptual error at data analysis only serves to indicate that data collection was misconducted due to a faulty (or even absent) hypothesis.

Regarding effect mediation, TL rather than TL genetic proxies may be preferred, since environmental exposures that are mediated by TL will not affect the germline genetic profile of an individual. In this context, TL would first be considered an outcome of an environmental exposure (a; if interest lies only in identifying factors associated with telomere length, the analysis would end at this point) and, then, investigated for association with a given outcome ( $e_1$ ) that is associated with the same risk factor (b); the attenuation of b when combining all relevant variables (as described above) can also be used to provide additional evidence for the mediation effect of TL. Importantly, given the robustness of genetic proxies (discussed in the previous sections), TL genetic proxies would still be of great value to estimate  $e_1$ . For the case of effect mediation, then, it can be of special usefulness to both measure TL and genotype TL genetic proxies. As an example of this combination, a genome-wide association of TL and bladder cancer used a multi-stage approach in which four single nucleotide polymorphisms were highly associated with blood leukocytes TL. Subsequently, these genetic variants were investigated for association with bladder cancer in a large case-control design, finding association with one of the variants, consistent with the association of this variant with TL. This variant was, then, investigated for the mediation effect of TL in its causal pathway (genetic variant  $\rightarrow$  TL  $\rightarrow$  bladder cancer), finding that TL was a significant mediator [18]. These findings confirm both the usefulness and the need of combining TL measurement and the use of TL genetic proxies, since the latter provides a robust estimate of  $e_1$  and, as shown by the study, genetic variants may have pleiotropic effects, which would have resulted in this study (if TL had been considered the only mediator of the genetic variant *a priori*), in over-estimating  $e_1$ .

### Causal Factor or Biomarker?

The two discussed approaches for studying telomere biology have an important theoretical limitation: both fail to address specificity. In

case of TL measurement, it is not necessarily true that TL status in the tissue where the measurement was taken is a satisfactory representation of the tissue of interest for the particular disease, indicating that TL measurement in accessible tissues should be considered, perhaps for most studies, a biomarker of disease risk/status rather than a factor itself implicated in disease etiology; in this regard, intra-individual TL variability has been evidenced to significantly exceed inter-individual variability (as reviewed elsewhere [13]). Regarding TL genetic proxies, the situation is somewhat the opposite: since cells in different compartments will suffer the effects of germ line genetic variants, TL in all these compartments will be associated with disease, although most likely just one or a few of these compartments are functionally implicated in disease causation. Although this might seem a minor concern, it has important implications for targeted therapies aiming at delaying telomere shortening, which need to take the cell compartment/tissue where TL plays an actual causal role in a given outcome into account (although TL can appropriately be used as a monitoring marker of a more systemic intervention that reduces both TL shortening in accessible tissues as well as disease risk, for example, without the need of identifying the cell compartment where TL is causal). In this context, neither TL measurement in an accessible tissue nor genotyping TL genetic proxies provides specificity. To overcome this, theoretical considerations of which cell compartments would be of greater importance to the outcome under investigation could be applied to studies using TL genetic proxies, but empirical evidence can be obtained, for example, from subjecting animals to an exposure of interest and measure TL of different cell compartments. Of great importance would be to measure TL in different adult stem cell compartments, since these are essential in tissue self-renewal maintenance and are protected from damages and stress due to their slow replicative rate and their protective microenvironment [32], which plausibly results in a very specific group of exposures being able to affect adult stem cells TL. Here, the issue of representativeness of TL measurement in an accessible source may be even more pronounced, thus requiring specific investigation. Evidently, TL genetic proxies cannot be used as biomarkers of disease status (only disease pre-disposition), since they are immutable (disregarding somatic mutations) during the lifespan.

### Final Remarks

It is clear that the use of genetic proxies is a valuable alternative to aid confidence to causal inference regarding the roles of TL in disease etiology, mainly due to the robustness of germline genetic profiles against reverse causation (due to always meet the temporality criteria, which is of special importance in one-time measurement designs such as cross-sectional studies) and confounding (assuming no or controlled population stratification). However, the dynamic nature of TL measurement is required for studying it as an outcome, which is required for studies of the roles of TL in mediating disease risk and of potential interventions aimed at telomere lengthening (or at least reducing shortening). For this end, TL genetic proxies are still useful, although robust causal inferences for the roles of environmental exposures on TL regulation ultimately depend on the use of an appropriate study design. In addition, the importance of an appropriate conceptual framework to guide both data collection and analysis (which is often underestimated [23]) was stressed, since the use of the most appropriate study design is frequently not feasible.

### References

1. Greider CW, Blackburn EH (1987) The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell 51: 887-898.

2. Greider CW, Blackburn EH (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* 43: 405-413.
3. Abbott A (2009) Chromosome protection scoops Nobel. *Nature* 461: 706-707.
4. de Lange T (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* 19: 2100-2110.
5. Hartwig FP, Collares T (2013) Telomere dysfunction and tumor suppression responses in dyskeratosis congenita: Balancing cancer and tissue renewal impairment. *Ageing Res Rev* 12: 642-652.
6. Deng Y, Chan SS, Chang S (2008) Telomere dysfunction and tumour suppression: the senescence connection. *Nat Rev Cancer* 8: 450-458.
7. Armanios M, Blackburn EH (2012) The telomere syndromes. *Nat Rev Genet* 13: 693-704.
8. Shay JW, Bacchetti S (1997) A survey of telomerase activity in human cancer. *Eur J Cancer* 33: 787-791.
9. Heaphy CM, Subhawong AP, Hong SM, Goggins MG, Montgomery EA, et al. (2011) Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. *Am J Pathol* 179: 1608-1615.
10. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646-674.
11. Eissenberg JC (2013) Telomeres, cancer & aging: live long & prosper? *Mo Med* 110: 11-16.
12. Armanios M (2013) Telomeres and age-related disease: how telomere biology informs clinical paradigms. *J Clin Invest* 123: 996-1002.
13. Prescott J, Wentzensen IM, Savage SA, De Vivo I (2012) Epidemiologic evidence for a role of telomere dysfunction in cancer etiology. *Mutat Res* 730: 75-84.
14. Mason PJ, Bessler M (2011) The genetics of dyskeratosis congenita. *Cancer Genet* 204: 635-645.
15. Mangino M, Richards JB, Soranzo N, Zhai G, Aviv A, et al. (2009) A genome-wide association study identifies a novel locus on chromosome 18q12.2 influencing white cell telomere length. *J Med Genet* 46: 451-454.
16. Codd V, Mangino M, van der Harst P, Braund PS, Kaiser M, et al. (2010) Common variants near TERC are associated with mean telomere length. *Nat Genet* 42: 197-199.
17. Levy D, Neuhausen SL, Hunt SC, Kimura M, Hwang SJ, et al. (2010) Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc Natl Acad Sci U S A* 107: 9293-9298.
18. Gu J, Chen M, Shete S, Amos CI, Kamat A, et al. (2011) A genome-wide association study identifies a locus on chromosome 14q21 as a predictor of leukocyte telomere length and as a marker of susceptibility for bladder cancer. *Cancer Prev Res (Phila)* 4: 514-521.
19. Prescott J, Kraft P, Chasman DI, Savage SA, Mirabello L, et al. (2011) Genome-wide association study of relative telomere length. *PLoS One* 6: e19635.
20. Mangino M, Hwang SJ, Spector TD, Hunt SC, Kimura M, et al. (2012) Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Hum Mol Genet* 21: 5385-5394.
21. Smith GD, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 32: 1-22.
22. Ebrahim S, Davey Smith G (2008) Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet* 123: 15-33.
23. Victora CG, Huttly SR, Fuchs SC, Olinto MT (1997) The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *Int J Epidemiol* 26: 224-227.
24. Shen Q, Zhao X, Yu L, Zhang Z, Zhou D, et al. (2012) Association of leukocyte telomere length with type 2 diabetes in mainland Chinese populations. *J Clin Endocrinol Metab* 97: 1371-1374.
25. Bjelakovic G, Beninati S, Pavlovic D, Kocic G, Jevtovic T, et al. (2007) Glucocorticoids and oxidative stress. *J Basic Clin Physiol Pharmacol* 18: 115-127.
26. Sampson MJ, Winterbone MS, Hughes JC, Dozio N, Hughes DA (2006) Monocyte telomere shortening and oxidative DNA damage in type 2 diabetes. *Diabetes Care* 29: 283-289.
27. Aviv A (2006) Telomeres and human somatic fitness. *J Gerontol A Biol Sci Med Sci* 61: 871-873.
28. Blackburn EH (2005) Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. *FEBS Lett* 579: 859-862.
29. Garcia-Calzon S, Gea A, Razquin C, Corella D, Lamuela-Raventos RM, et al. (2013) Longitudinal association of telomere length and obesity indices in an intervention study with a Mediterranean diet: the PREDIMED-NAVARRA trial. *Int J Obes (Lond)*.
30. Khan S, Chuturgoon AA, Naidoo DP (2012) Telomeres and atherosclerosis. *Cardiovasc J Afr* 23: 563-571.
31. Kiecolt-Glaser JK, Epel ES, Belury MA, Andridge R, Lin J, et al. (2013) Omega-3 fatty acids, oxidative stress, and leukocyte telomere length: A randomized controlled trial. *Brain Behav Immun* 28: 16-24.
32. Morrison SJ, Spradling AC (2008) Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* 132: 598-611.