HEALTH AND RETIREMENT STUDY

2008 Telomere Length Data

Version 1.0, December 2013
(Sensitive Health Data)

Data Description and Usage

If there are any questions about this data set and its use, please contact the HRS Help Desk at hrsquestions@umich.edu.

Contents

1. Introduction

Average telomere length was assayed using quantitative PCR (qPCR), adapted from the published original method by Cawthon [1], by comparing telomere sequence copy number in each patient's sample (T) to a single-copy gene copy number (S). The resulting T/S ratio is proportional to mean telomere length [1, 2]. The 2008 Telomere Data include average telomere length data from 5808 HRS respondents who consented and provided a saliva sample during the 2008 interview wave. Assays were performed by Telome Health (Telomere Diagnostics, http://www.telomehealth.com/)

1.   Cawthon, R.M., *Telomere measurement by quantitative PCR.* Nucleic Acids Res, 2002. 30(10): p. e47.
2.   Aviv, A., et al., *Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR.* Nucleic Acids Res, 2011.

1a. Background

Telomeres consist of repeated DNA that create protective caps at the ends of chromosomes. Telomeres protect chromosomes from shortening and damage that may ultimately lead to cellular death and loss of health. Telomeres shorten over time, and this shortening is considered both a marker of cell aging as well as a causal factor in cell aging.

Short telomeres impair the ability of cells to divide properly. When the DNA in a cell is unable to replicate properly, the cell either undergoes cell death, or functions poorly.

Studies have shown associations between shorter telomere length and various types of cardiovascular disease (e.g., stroke, heart attacks), cancer, and diabetes. Shorter telomeres have also been associated with osteoporosis, cognitive function, dementia, depression, and inflammatory diseases like arthritis.

See the Telomere Diagnostics webpage for more information: http://www.telomehealth.com/

Sample Selection for the Enhanced Face-to-Face Interview

In 2006, HRS initiated what is referred to as an Enhanced Face-to-Face Interview (EFTF). In addition to the core interview, the EFTF interview includes a set of physical performance tests, anthropometric measurements, blood and saliva samples, and self-administered questionnaire on psychosocial topics.

Approximately fifty percent of households with at least one living respondent were selected for the EFTF interview across all primary sampling units (PSUs). A random one half of the 2006 sample was preselected to the EFTF interview. The other half was selected in 2008. The sample was selected at the household-level to ensure that the same request was made to both members of a household. New spouses of respondents flagged to complete an EFTF interview were also asked to do so. Thus, in coupled households, both members of the couple were selected. Some respondents who were selected for the enhanced face-to-face sample were not asked to complete the physical measures or biomarkers. This group included respondents who a) needed to be interviewed by proxy, b) resided in a nursing home, or c) declined a face-to-face interview but agreed to be interviewed by telephone.

The preload variable that identifies the EFTF sample is LX090_R for 2008. A value of 3 indicates that the respondent was assigned to the EFTF sample for that wave. The variable EFTFASSIGN in the Cross-wave Tracker file indicates the respondents' permanent assignment for enhanced face-to-face rotation for 2006 and beyond.

Consent Procedures

Prior to conducting the saliva collection, a consent form was administered by the interviewer. Respondents were asked to read and sign the form. Respondents who did not sign the consent form were not asked to complete the collection. Separate consent forms were administered for the saliva and blood samples. Each form was introduced just prior to the collection or measure that it covered. After obtaining consent for a given component, the interviewer described the procedures in detail to the respondent and demonstrated how the collection would be conducted.

DNA Collection and Extraction

As noted above, the saliva collection was intended for all of those who were available for the EFTF interview. See general sampling description (above). The saliva consent rate in 2008 was 85%. The completion rate, conditional on consent, was 99%. The overall completion rate was 84%. A saliva sample was obtained directly using an Oragene Collection Kit. Saliva samples were immediately sent to a central laboratory where they were processed. DNA was extracted and stored.

1b. Analytic Methods and Data Quality

Participant DNA samples

DNA samples were sent to Telome Health in 96 well plates. Well positions F12-H12 are empty for plates 1-63. Plate 64 is a partial plate with wells B8-H12 empty. All DNA samples were stored in their original plates in a -80 freezer upon arrival. DNA samples were thawed on ice to make diluted working plates. The diluted working plates were stored at 4°C and assayed within 1 week. The original sample plates were put back to a -80 freezer immediately after dilution. Based on the DNA concentrations provided by HRS study, the source DNA was diluted with low EDTA TE resulting in DNA concentrations within the reference standard curve in working plates. Specifically, a 10-fold dilution was made for plates 2-57, plates 62 and 63, an 8-fold dilution made for plate 1, a 4-fold dilution was made for plate 58, 5-fold dilution made with plates 59 and 60 and 3-fold dilution was made for plate 61.

Primers

Primers for the telomere PCR (T runs) are *tel1b* [5'-CGGTTT(GTTTGG)$_5$GTT-3'], used at a final concentration of 100 nM, and *tel2b* [5'-GGCTTG(CCTTAC)$_5$CCT-3'], used at a final concentration of 900 nM. The primers for the single-copy gene (human beta-globin) PCR (S runs) are *hbg1* [5' GCTTCTGACACAACTGTGTTCACTAGC-3'], used at a final concentration of 300 nM, and *hbg2* [5'-CACCAACTTCATCCACGTTCACC-3'], used at a final concentration of 700 nM.

All primers were purchased from IDT (*www.idtdna.com) in HPLC purified form.*

Standard reference DNA

Genomic DNA from pooled 100 male donors (Mosaic human gDNA, M, from Aldevron, cat# 5080-25) was used as the standard reference. A 3- fold serial dilution was made to create a 6-point standard curve, containing 5, 1.6667, 0.5556, 0.1852, 0.06173 and 0.02058 ng of DNA respectively in each reaction tube. Each assay plate contains quadruplicate wells for each of the standard curve concentrations.

Reagents

- E. coli DNA stock (480ng/ul)

Nuclease-free ddH$_2$O (Invitrogen, cat# 10977-015) was used to resuspend lyophilized E. coli DNA (Sigma-Aldrich, cat# D2001) to 480 ng/ul. The resuspended DNA was vortexed at room temperature for 15 minutes, stored at 4°C overnight and then stored at -20 °C until further use.

- Denaturation buffer (25 mM Tris-HCl, pH 8.4; 62.5 mM KCl, 3.8 ng/ul of E. coli DNA)

Denaturation buffer was made by using 10XPCR buffer (200 mM Tris-HCl, pH 8.4; 500 mM KCl) supplied in the Platinum Taq kit (Invitrogen, cat #10966-083) and 480 ng/ul of E coli DNA stock.

Denaturation buffer for the entire HRS study was made as a single batch and stored as individual aliquots for a set of 3 runs (3 T runs and 3 S runs) at -20°C.

- U, T and S mixes

U mix contains Tris.HCl, PH 8.4; KCl, MgCl2, dNTPs and DMSO
U mix was made from 10XPCR buffer and 50 mM MgCl$_2$ (from Platinum Taq kit), 100 mM dNTP (Roche Applied Science, cat # 03622614001) and DMSO (Sigma-Aldrich, cat# 154938).
T mix contains tel 1b and tel 2b primers
S mix contains hbg1and hbg2 primers and MgCl$_2$

U, T and S mixes were made as a single batch and stored as individual aliquots for each run at -20°C.

- Sybr Green I

Working stocks of Sybr Green I (10X) was made from 10,000 X in DMSO (Invitrogen, cat# S7585) with ddH$_2$O , aliquoted in black 1.5mL sterilized tubes for each pair of T and S runs, stored at -20 °C and used within one week after preparation.

- Platinum Taq polymerase

Platinum Taq polymerase was purchased from Invitrogen, aliquoted into single usage (one run) and stored at -20°C.

The final reaction mix contains 20 mM Tris-HCl, pH 8.4; 50 mM KCl; 200 µM each dNTP; 1% DMSO; 0.5x SRBR Green I; 16.5 ng E. coli DNA per reaction; 0.4 Units of Platinum Taq DNA polymerase per 11 microliter reaction.

Experimental procedures

40 ul of denaturation buffer and 10 ul of DNA from diluted working DNA plates were pipetted into a 96 well plate (Biorad Hard Shell 96-well plates, cat # HSP9901) by a MicroLab NIMBUS liquid handler (Hamilton Robotics). DNA samples were heated at 95 °C for 10 minutes in a Thermal cycler (Biorad Model C1000, cat # 185-1096) and kept on ice afterwards. Similarly, 3-fold serial dilution of the reference standard DNA was made using denaturation buffer, heated at 95 °C for 10 minutes in the Thermo cycler and kept on ice.

A complete master mix was made by combining U mix, Sybr Green, Taq polymerase and T mix (for T run), or S mix (for S run). 5.5 ul of the complete master mix and 5.5 ul of the denatured DNA were pipetted into each well of the 384 well PCR plate (Roche Applied Science, cat # 0510243001) by the MicroLab NIMBUS liquid handler. Each denatured participant DNA sample was pipetted into triplicate wells and each standard reference DNA was pipetted into quadruplicate wells. The 384 well plate was then put into Roche LightCycler 480 for qPCR.

The above mentioned experimental procedure was performed for plates 2-13, 15, 16, 18, 20-25, 27-48. For the remaining plates, a slightly modified procedure that streamlined the process was used. Specifically, 68 ul of denaturation buffer and 17 ul of DNA from diluted working DNA plates were pipetted into a 96 well plate (Biorad Hard Shell 96-well plates, cat # HSP9901) by a MicroLab NIMBUS liquid handler (Hamilton Robotics). DNA samples were heated at 95 °C for 10 minutes in a Thermal cycler (Biorad Model C1000, cat # 185-1096) and kept on ice afterwards. Similarly, 3-fold serial dilution of the reference standard DNA was made using denaturation buffer, heated at 95 °C for 10 minutes in the Thermo cycler and kept on ice. The denatured sample DNA plate and the reference standard DNA were used for the first pair of T and S runs, were then denatured again before being used for the second pair of T and S runs. Comparison of these two experimental procedures with 93 HRS sample showed an average CV of 2.4% between these 2 procedures.

The Thermal cycling profiles are:

T run: denature at 96°C for 1 second, anneal/extend at 54°C for 60 seconds, with fluorescence data collection, 30 cycles.

S run: denature at 95°C for 15 seconds, anneal at 58°C for 1 second, extend at 72°C for 20 seconds, 8 cycles; followed by denature at 96°C for 1 second, anneal at 58°C for 1 second, extend at 72°C for 20 seconds, hold at 83°C for 5 seconds with data collection, 35 cycles.

Sample exclusion

A total of 5916 participant samples (93 samples from plates 1-63 and 57 samples from plate 64) from the HRS cohort were tested. Of these participant samples, 5257 samples (88.9%) passed all QC criteria the first pass. Additionally, due a systematic error in the A1 position of the liquid handler, all samples in the A1 position from all plates were retested. This resulted in 719

retesting samples. In total, 76 samples failed QC criteria in the retesting, resulting in 5840 passed samples with an overall pass rate of 98.72%. Of the failed samples, 21 had no detectable values in the assays. Furthermore, many failed samples were discolored or viscous. After completion of analysis, an additional 2 samples we dropped because no T/S ratio was reported, 4 were determined to be unintended duplicates, and 26 had sample identity issues. The final analytic file includes 5808 samples for a final success rate of 98.17%.

QC criteria
The following QC criteria were applied to HRS study samples:

- Standard curve

In order to reproducibly generate T/S ratios between assay runs, standards curves are generated for both T and S. For this purpose, three-fold serial dilutions of known T and S starting concentrations are generated in quadruplicate qPCR wells on each qPCR plate. qPCR is then performed in order to obtain corresponding crossing point (C_p) measurements. Standard curves are generated by modeling the relationship between the observed C_p measurements and known concentration levels using a model developed by Telome Health.

Amplification efficiency of the standard curves
The accepted criteria of PCR amplification efficiency is >75% for T runs and >80% for S runs.

Assessments of the linearity of each standard curve
Assessments of the linearity of each standard curve for T and S were performed using a statistical test of lack-of-fit from the linear regression model with replicate measurements at each standard concentration. Departures were considered significant if the resulting p-value was less than 0.01. In such cases, departure from linearity are said to be statistically significant. When departures from linearity were detected, attempts were made to eliminate the lowest concentrations first. If that failed, then attempts were made to identify specific standards that seem to be abnormal relative to the rest of the data and such standard points were deleted from the curve.

Individual outlier detection in standard curve and participant samples
Studentized deleted residual was used to detect individual outlier wells from standard curves.

Studentized outlier removal test was used to remove wells with unusual Cp callings in the triplicate wells for participant samples.

- CV acceptance criteria

CV(coefficient of variation) for each sample was calculated based on three runs (three pairs of T and S runs) for plates 2-9, 11 and 13 and based on two runs for plates 1, 10, 13-64. Samples that had smaller than 12.5% CV were considered as pass and samples with greater than 12.5% CV were reassayed.

Comparison of CVs of 10 quality control DNA samples from 10 plates each with 3 runs showed that 3 runs compared to 2 runs do not reduce CVs. Therefore, all subsequently plates were assayed with 2 runs.

1c. Data Summary

Analytic performances:

- PCR efficiencies

Summary of T and S run efficiencies is listed in Table 1

Table 1	T runs	S runs
average	0.884	0.937
stdev	0.032	0.033

The maximum amplification efficiency for T is 98% and the minimum amplification efficiency observed was 77.3%.

The maximum efficiency for S is 100% and the minimum amplification efficiency observed was 82.4%.

- Control DNA samples

Genomic DNA from 3 cancer cell lines was included in each run to assess the analytic performance of each PCR run. Table 2 summaries the results of these control DNA samples

Table 2	mean T/S ratio	STDEV	CV
control 1	0.528	0.030	5.6%
control 2	1.147	0.072	6.3%
control 3	2.317	0.081	3.5%

2. Obtaining the Data

2a. How to Apply

1. Visit the <u>HRS User Registration/File Download Web site</u>. If you do not already have a username and password, you must register in order to enter the site.
2. Identify the desired data set (Telomere 2008, in this case) under the *Requesting HRS Special Access Files* heading.
3. Download and complete the Data Use Agreement (DUA) in hardcopy format (See Appendix for sample).
4. Send a signed copy of the agreement either:
    a. via surface mail to:
        Health and Retirement Study
        DUA Review Committee
        426 Thompson Street, Room 3050 ISR
        Ann Arbor, Michigan 48104-2321
    b. or fax it to: (734) 615-5433.

2b. Approval Process

The Health and Retirement Study will review your request, and based on the information that you provided when you registered at the <u>HRS User Registration/File Download Web site</u>, verify your identity and institutional affiliation. Once this authentication process has been completed to our satisfaction, we will authorize access to the *2008 Telomere* data set. We will communicate with you at the email address that you provided when you registered at the download site.

2c. How to Download

Once you receive your approval notification, login to the HRS User Registration/File Download Web site. In the HRS Special Access Files box on the right hand side of the page you will now see links to the *2008 Telomere Data*. Click on the appropriate link and follow instructions to download the Telomere files.

2c. Publications Based on Restricted Data

Your sensitive health data agreement specifies that you will inform HRS of any papers, publications, or presentations based on this restricted data set. Please send a copy of any such publication, with a bibliographical reference (if appropriate), in PDF format as an electronic mail attachment to the HRS Help Desk at http://hrsonline.isr.umich.edu/help "Attn: Papers and Publications" on the subject line. As an alternative, you may send publications via the postal service to this address:

```
Health and Retirement Study
Attn: Papers and Publications
The Institute for Social Research, Room 3050
P.O. Box 1248
Ann Arbor, Michigan  48106-1248
```

The following citation should be included in any research reports, papers, or publications based on these data:

In text: "The HRS *2008 Telomere* data set is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and was conducted by the University of Michigan."

In references: "Health and Retirement Study, *2008 Telomere* data set. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIAU01AG009740). Ann Arbor, MI, (2013)."

3. Setup

3a. Distribution Set

The HRS *2008 Telomere Length* data set is packaged for distribution in a .ZIP file, `telomere08.zip`. In order to keep the contents secure, the .ZIP file has been encrypted using WinZIP 256 bit AES encryption. A WinZIP self-extracting .exe file is distributed for users working in a Microsoft Windows environment. Unix, Linux and Macintosh users receive a standard encrypted .ZIP file. Extract the data file, the program statement file that matches your analysis environment, the data description (this file), and the codebook file. If you require a special file format or experience system problems, please contact the HRS Help Desk (http://hrsonline.isr.umich.edu/help).

3a-1. Windows Environment

Run the self-extracting file, `telomere08.exe`, from the command line or from Windows Explorer. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output from the self-extracting archive file will be the contents of compressed file `telomere08.zip`, which contains the files listed in Table 1. At a minimum you will need to save the data file, one set of program statement files, and the codebook file. You will need approximately 2 MB of free space on your storage device to store all the files.

3a-2. UNIX/Linux Environment

Use the .ZIP file decompression software installed on your system, (e.g. **7z, gunzip**) to decrypt and decompress `telomere08.zip`. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output from `telomere08.zip` will be the files listed in Table 1. At a minimum you will need to save the data file, one set of program statement files, and the codebook file. You will need approximately 2 MB of free space on your storage device to store all the files.

3a-3. Macintosh OS X Environment (10.4.x and above)

Use **stuffit** to decrypt and decompress `telomere08.zip`. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output from `telomere08.zip`, will be the files listed in Table 1. At a minimum you will need to save the data file, one set of program statement files, and the codebook file. You will need approximately 2 MB of free space on your storage device to store all the files.

Table 1: Contents of Distribution Package

Directory[1]	File	Type
`c:\telo\data\`	`TELOMERESA_R.da`	`Data files (ASCII text)`
`c:\telo\docs\`	`TELOMERESA_R.txt`	`Codebook files (ASCII text)`
`c:\telo\sas\`	`TELOMERESA_R.sas`	`SAS program statements`
`c:\telo\spss\`	`TELOMERESA_R.sps`	`SPSS program statements`
`c:\telo\stata\`	`TELOMERESA_R.dct/` `TELOMERESA_R.do`	`Stata dictionary and "do" files`

3b. Program Statements

The *2008 Telomere* data set includes SPSS, SAS or Stata program statements that are used to read an ASCII data file. Files containing SPSS statements are named with an .sps extension, those with SAS statements with an .sas extension, and those with Stata statements with .do and .dct extensions. The files are keyed on [HHID and PN].

3b-1. Using the Files with SAS

To create a SAS system file from the *2008 Telomere* data set, two file types must be present: the SAS program statement file (`TELOMERESA_R.sas`) and the ASCII data file (`TELOMERESA_R.da`). To create a SAS system file, load the file into the SAS Program Editor. If the program statement file is located in "c:\telo\sas" and the data file is located in "c:\telo\data", you can run the file as is. A SAS system will be saved to directory "c:\telo\sas". If the files are not located in the specified directories, you will need to edit the `TELOMERESA_R.sas` file to reflect the proper path names prior to running the file.

3b-2. Using the Files with SPSS

To create an SPSS system file from the *2008 Telomere Length* data set, two file types must be present: the SPSS program statement file (`TELOMERESA_R.sps`) program statement file and the ASCII data file (`TELOMERESA_R.da`). To create an SPSS system file, open the program statement file in SPSS as an SPSS Syntax File. If the program statement file is located in "c:\telo\spss" and the data file is located in "c:\telo\data", you can run the file as is. An SPSS system file will be saved to directory "c:\telo\spss". If the files are not located in the specified directories, you will need to edit the `TELOMERESA_R.sps` file to reflect the proper path names prior to running the file.

3b-3. Using the Files with Stata

To create a Stata system file from the *2008 Telomere Length* data set, three file types must be present: the Stata dictionary file (`TELOMERESA_R.dct`), the Stata do- file (`TELOMERESA_R.do`), and the ASCII data file (`TELOMERESA_R.da`) containing the raw data for Stata to read. To create a Stata system file (`TELOMERESA_R.dta`), load the do-file into Stata and then submit it. If the *.do and .dct files are located in "c:\telo\stata" and the data file is located in "c:\telo\data", you can run the .do file as is. If the files are not located in these

---

[1] While a fixed file structure is not required for using HRS files, we have traditionally suggested a directory structure for the Windows environment. By using this directory structure, you will not have to change the path name references in your data descriptor files. If you use a different structure, just change the directory references in the program files.

directories before you run the files you must edit the `TELOMERESA_R.do` and `TELOMERESA_R.dct` files to reflect the proper path names prior to running the do-file

4. Data Set Content

This version of the *2008 Telomere* data set includes average telomere length data from 5808 HRS respondents who consented and provided a saliva sample during the 2008 interview wave. It has 5808 records, which are uniquely identified by HHID (household ID) and PN (person number). See Section 1 of this document for additional information.

5. If You Need to Know More

This document is intended to serve as a brief overview of the *2008 Telomere* data set. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

5a. HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access public data or to find out more about sensitive health data products and procedures, visit the HRS sensitive health Web site.

5b. Contact Information

If you need to contact us, you may do so by one of the methods listed below.

```
Internet: Help Desk at the HRS Web site

E-mail: hrsquestions@umich.edu

Postal Service:
     Health and Retirement Study
     The Institute for Social Research
     426 Thompson Street, 3050 ISR
     Ann Arbor, Michigan  48104

FAX: (734) 615-5433
```

Appendix: Health and Retirement Study Sensitive Health Data Use Agreement

Data Set Requested: HRS 2008 Telomere

By signing this form and obtaining the requested data set from HRS, User agrees:

1. To use the data set solely for statistical reporting and analysis.

2. Not to share these data with, or provide copies of these data to, any other person or organization. Note: Each research assistant/associate, graduate student or undergraduate student, or other individual working on the research project must sign and submit their own form.

3. To return or destroy the data set, and any derivative data files, upon request from HRS.

4. To make no attempt to link this data set with individually identifiable records from any source, or in any other way attempt to identify the persons in this or other HRS datasets.

5. That if the identity of any person or establishment in this data set is inadvertently discovered, then (a) no use will be made of this knowledge, (b) the Director of HRS will be advised of this incident immediately, (c) the information that would identify any individual or establishment will be safeguarded or destroyed, as requested by HRS, and (d) no one else will be informed of the discovered identity.

6. To employ the following guidelines when producing tabulations for distribution:
    - Magnitude Data: Ensure that no cells/strata with n < 5 are produced.
    - Frequency Data: Apply a marginal threshold of n >= 5 and cell threshold of n >= 5 to all tabulations.
    - Protecting against complementary disclosure: Additional cells may be suppressed, i.e., complementary disclosure, to make sure the primary suppressions cannot be derived by subtraction from published marginal totals.

7. Aggregate statistical summaries of the data and analyses (frequency tabulations, magnitude tabulations, means, variances, regression coefficients, and correlation coefficients): These are approved under this agreement and may be freely published by the User, subject to the provisions above.

8. To cite HRS as the data source in any publications or research based upon these data, and to provide a copy of any publications to the HRS. The following citation should be included in any research reports, papers, or publications based on these data:

    In text: "The HRS 2008 Telomere data set is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and was conducted by the University of Michigan."

    In references: "Health and Retirement Study, 2008 Telomere data set. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (2013)."

9. To maintain a secure computing environment for storage and use of this data set and any data sets derived from it.

10. To hold harmless and indemnify HRS and the University of Michigan, its agents and employees, for any claims of breaches of confidentiality arising out of my research, defined as failure to abide by any section of this agreement or any accidental or intentional violation of privacy of any contributor to any HRS data resource.