

Mean Cumulative Count under Left-Truncated Data in the Presence of Competing Risk: a CCSS Study

Grace Zhou
Biostatistician | Department of Biostatistics
2025-2-20



Overview

Background

Conventional methods in survival analysis

Method

- Geskus 2011: weight method for left-truncated data with competing risks

Application

 Mean Cumulative Count (MCC) in a Childhood Cancer Survivor Study (CCSS)

Conclusion



Survival analysis

- Survival function
 - o Definition: probability that an individual survives to time t
 - Method: Kaplan-Meier estimates (Kaplan & Meier 1958)
- Hazard function
 - Definition: instantaneous rate of the occurrence of the event of interest in subjects currently at risk
 - Method: Cox proportional hazards regression model (Cox 1972)



- Survival analysis in the presence of competing risk
 - Competing risk
 - An event precludes the occurrence of the primary event of interest
 - Hazard function
 - o Cause-specific hazard function (Prentice et al. 1979)

$$\lambda_k^{cs}(t) = \lim_{\Delta t \to 0} \frac{\operatorname{Prob}(t \leq T < t + \Delta t, D = k)}{\Delta t},$$

Subdistribution hazard function (Fine & Gray 1999)

$$\lambda_k^{sd}(t) = \lim_{\Delta t \to 0} \frac{\operatorname{Prob}(t \leq T < t + \Delta t, D = k | T \geq t \cup (T < t \cap K \neq k))}{\Delta t}.$$

event-free + who have previously experienced a competing risk

- Hazard model
 - Cause-specific hazard model
 - ✓ R: survival::coxph(Surv(time, status==1)~covariates,data)
 - Fine-Gray subdistribution hazard model
 - ✓ R: cmprsk::crr(ftime, fstatus, covariates)

event-free



- Survival analysis in the presence of competing risk
 - Competing risk
 - An event precludes the occurrence of the primary event of interest
 - Cumulative incidence function (CIF)
 - o The incidence of a specific event occurring by a certain time, considering competing risks
 - $\circ CIF_{k(t)} = \Pr(T \le t, D = k) = \int_0^t \lambda_k(u)S(u)du$
 - \circ S(u): probability of not experiencing any event by time u, including both event of interest and competing risk event
 - o a single cause-specific hazard function is insufficient
 - o Fine-Gray subdistribution hazard can directly model CIF
 - R: cmprsk::cuminc(ftime,fstatus,group) or tidycmprsk::cuminc(Surv(ftime,fstatus)~group)



- Survival analysis with competing risk & left-truncation
 - Left-truncation
 - An unknown subset of subjects failed before a certain time and the subjects didn't get into the study (Harrell 2015)
 - Hazard model
 - Cause-specific hazard model is valid by using the counting process (Therneau & Grambsch 2000)
 - √ R: survival::coxph(Surv(tstart,tstop,status==1)~covariates,data)
 - o Fine-Gray subdistribution hazard model cannot handle left-truncated data directly
 - CIF
 - Cannot be determined directly as Fine-Gray subdistribution hazard is available



Cause-Specific Cumulative Incidence Estimation and the Fine and Gray Model Under Both Left Truncation and Right Censoring

Ronald B. Geskus

Academic Medical Center, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Amsterdam Health Service, Nieuwe Achtergracht 100, 1018 WT Amsterdam, The Netherlands email: statistics@inter.nl.net

Summary. The standard estimator for the cause-specific cumulative incidence function in a competing risks setting with left truncated and/or right censored data can be written in two alternative forms. One is a weighted empirical cumulative distribution function and the other a product-limit estimator. This equivalence suggests an alternative view of the analysis of time-to-event data with left truncation and right censoring: individuals who are still at risk or experienced an earlier competing event receive weights from the censoring and truncation mechanisms. As a consequence, inference on the cumulative scale can be performed using weighted versions of standard procedures. This holds for estimation of the cause-specific cumulative incidence function as well as for estimation of the regression parameters in the Fine and Gray proportional subdistribution hazards model. We show that, with the appropriate filtration, a martingale property holds that allows deriving asymptotic results for the proportional subdistribution hazards model in the same way as for the standard Cox proportional hazards model. Estimation of the cause-specific cumulative incidence function and regression on the subdistribution hazard can be performed using standard software for survival analysis if the software allows for inclusion of time-dependent weights. We show the implementation in the R statistical package. The proportional subdistribution hazards model is used to investigate the effect of calendar period as a deterministic external time varying covariate, which can be seen as a special case of left truncation, on AIDS related and non-AIDS related cumulative mortality.

KEY WORDS: Competing risks; Inverse probability weight; Subdistribution hazard; Survival analysis.



Geskus 2011

- Individuals who experienced competing risk receive time-dependent weights from censoring and truncation mechanisms
 - Censoring mechanism

Let $\widehat{\overline{G}}$ denote the product-limit statistic, which is obtained by reversing the role of T and C:

$$\widehat{\overline{G}}(t) = \prod_{j:c_{(j)} \le t} \left(1 - \frac{m_j}{r(c_{(j)})} \right). \tag{5}$$

Truncation mechanism

$$\widehat{H}(t) = \prod_{-l_{(j)} < -t} \left(1 - \frac{w_j}{r(l_{(j)})} \right) = \prod_{l_{(j)} > t} \left(1 - \frac{w_j}{r(l_{(j)})} \right). \quad (6)$$

Weight

$$\omega_l(t_{(i)}) = \begin{cases}
\frac{\widehat{\overline{G}}(t_{(i)} -)}{\widehat{\overline{G}}(t_{(j)} -)} \frac{\widehat{H}(t_{(i)} -)}{\widehat{H}(t_{(j)} -)} & \text{if } l \text{ had competing event} \\
0 & \text{otherwise.}
\end{cases}$$
(11)

Note that $\widehat{\overline{G}}$ and \widehat{H} are only evaluated at the event times (of any type) and at the censoring times in \widehat{N} . As long as event times are separate from censoring and left entry times, we have $\widehat{\overline{G}}(t_{(i)}-)=\widehat{\overline{G}}(t_{(i)})$ and $\widehat{H}(t_{(i)}-)=\widehat{H}(t_{(i)})$.



Table 1
Three example individuals with different type of end point

id	Tstart	Tstop	stat	weight.cens	weight.trunc
1	0.25460	0.63644	censored		_
2	0.00000	0.64358	type 1	event	
3	0.08005	0.25615	type 2	competing risk	
Wi	th weights	S:			
1	$0.254\overline{60}$	0.63644	censored	1.00000	1.00000
2	0.00000	0.64358	type 1	1.00000	1.00000
3	0.08005	0.25615	type 2	1.00000	1.00000
3	0.25615	0.31778	type 2	1.00000	1.02941
3	0.31778	0.37693	type 2	1.00000	1.02941
3	0.37693	0.38928	type 2	1.00000	1.02941
3	0.38928	0.46029	type 2	1.00000	1.02941
3	0.46029	0.50979	type 2	1.00000	1.02941
3	0.50979	0.64358	type 2	0.67849	1.07230
3	0.64358	0.64724	type 2	0.67849	1.07230

event observed time

Cause-specific cumulative incidence in R practice

- wData=crprep2(Tstop="tstop", status="status", data=data, trans = 1, cens = 0,
 Tstart="tstart", id="id", shorten=FALSE
- Fit=survival::surfit(Surv(tstart,tstop,status==1)~Group, wData, weight=weight.cens*weight.trunc)
- CIF=1-summary(Fit)\$Surv

- •Proportional subdistribution (Fine-Gray) hazard model in R practice
 - wData=crprep2(Tstop="tstop", status="status", data=data, trans = 1, cens = 0,
 Tstart="tstart", id="id", shorten=FALSE)
 - survival::coxph(Surv(tstart,tstop,status==1)~covariates, wData, weight=weight.cens*weight.trunc)



- Example of weight data
 - Step 0: Create a dummy data

id	tstart	tstop	status
1	1	2	2
2	0	3	0
3	2	4	1
4	1	5	1



- Example of weight data
 - Step 1: calculate product-limit time-to-censoring distribution
 survival::survfit(Surv(tstart,tstop+ifelse(status==0, prec, 0),status==0)~1)

Time	n.event	survival
2	0	1
3	0	1
4	1	0.667
5	0	0.667



- Example of weight data
 - Step 2: calculate product-limit time-entry distribution at tsurvival::survfit(Surv(-tstop,-(tstart+2*prec),status==1)~1)

Time	n.event	survival
-2	1	0.667
-3	0	1
-4	0	1
-5	0	1



- Example of weight data
 - Step 3: prepare weight data
 mstate:::create.wData.omega(tstart, tstop, status, id, failcode=1, cens=0, stratum=1)

id	Tstart	Tstop	status
1	1	2	2
1	2	4	2
1	4	5	2
2	0	3	0
3	2	4	1
4	1	5	1



- Example of weight data
 - Step 4: assign weight
 - o data.weight\$weight.cens=summary(Gt,times=data.weight\$Tstop-prec)\$surv
 - o data.weight\$weight.trunc=summary(Ht,times=-data.weight\$Tstop+prec)\$surv

id	Tstart	Tstop	Weight.cens	Weight.trunc
1	1	2	1	0.667
1	2	4	0.667	1
1	4	5	0.667	1
2	0	3	1	1
3	2	4	0.667	1
4	1	5	0.667	1



- Example of weight data
 - Step 5: Finalize weight data

```
data.weight %>%
    group_by(id) %>%
    mutate(weight.cens=ifelse(status==2, weight.cens/first(weight.cens),1),
        weight.trunc=ifelse(status==2, weight.trunc/first(weight.trunc),1)) %>%
    ungroup()
```

id	Tstart	Tstop	Weight.cens	Weight.trunc
1	1	2	1	1
1	2	4	0.667	1.5
1	4	5	0.667	1.5
2	0	3	1	1
3	2	4	1	1
4	1	5	1	1

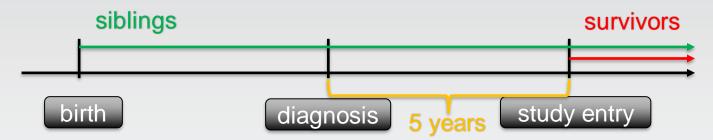


- Example of weight data
 - data <- data.frame(id=c(1,2,3,4), tstart=c(1,0,2,1), tstop=c(2,3,4,5), status=c(2,0,1,1))
 - crprep2(Tstop=data\$tstop,status=data\$status,id=data\$id,Tstart=data\$tstart,shorten=FA
 LSE)

id	Tstart	Tstop	status	weight.cens	weight.trunc	count	failcode
1	1	2	2	1.0000000	1.0	1	1
1	2	4	2	0.6666667	1.5	2	1
1	4	5	2	0.6666667	1.5	3	1
2	0	3	0	1.0000000	1.0	1	1
3	2	4	1	1.0000000	1.0	1	1
4	1	5	1	1.0000000	1.0	1	1



- Description of a Childhood Cancer Survivor Study (CCSS) study
 - Study: A CCSS study of Ewing Sarcoma (EWS)
 - Aim: Compare MCC of Grade 3-5 Musculoskeletal (MSK) Chronical Health Condition
 (CHC) between EWS survivors and siblings



- Left-truncation: Survivors who died before study entry are not included
- Competing risk: Death not due to MSK conditions
- Time scale: Time since birth



What is Mean Cumulative Count (MCC)



American Journal of Epidemiology

© The Author 2015. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Vol. 181, No. 7 DOI: 10.1093/aje/kwu289 Advance Access publication: February 17, 2015

Practice of Epidemiology

Estimating the Burden of Recurrent Events in the Presence of Competing Risks: The Method of Mean Cumulative Count

Huiru Dong, Leslie L. Robison, Wendy M. Leisenring, Leah J. Martin, Gregory T. Armstrong, and Yutaka Yasui*

* Correspondence to Dr. Yutaka Yasui, School of Public Health, University of Alberta, 3-381 Edmonton Clinic Health Academy, Edmonton, Alberta T6G 1C9, Canada (e-mail: yyasui@ualberta.ca).

Initially submitted May 1, 2014; accepted for publication September 19, 2014.

Cumulative incidence has been widely used to estimate the cumulative probability of developing an event of interest by a given time, in the presence of competing risks. When it is of interest to measure the total burden of recurrent events in a population, however, the cumulative incidence method is not appropriate because it considers only the first occurrence of the event of interest for each individual in the analysis: Subsequent occurrences are not included. Here, we discuss a straightforward and intuitive method termed "mean cumulative count," which reflects a summarization of all events that occur in the population by a given time, not just the first event for each subject. We explore the mathematical relationship between mean cumulative count and cumulative incidence. Detailed calculation of mean cumulative count is described by using a simple hypothetical example, and the computation code with an illustrative example is provided. Using follow-up data from January 1975 to August 2009 collected in the Childhood Cancer Survivor Study, we show applications of mean cumulative count and cumulative incidence for the outcome of subsequent neoplasms to demonstrate different but complementary information obtained from the 2 approaches and the specific utility of the former.

cumulative incidence; disease burden; mean cumulative count; recurrent events

How to obtain MCC?

$$MCC(t) = \sum_{p=1}^{m} CumI_p(t)$$

where $CumI_p(t)$ represents the cumulative incidence for the pth (p = 1, 2, ..., m) occurrence of the event of interest by time t.

Computation code

https://ccss.stjude.org/resourcetools.



Limitation

Recurrent events are assumed to be independent Finding cures. Saving children.



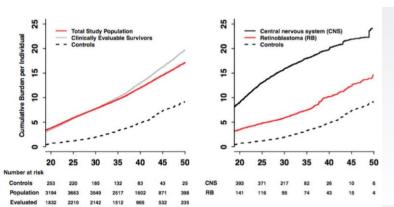
•When to apply MCC?

Lancet. 2017 Dec 9;390(10112):2569-2582. doi: 10.1016/S0140-6736(17)31610-0. Epub 2017 Sep 8.

The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE)

Nickhill Bhakta ¹, Qi Liu ², Kirsten K Ness ³, Malek Baassiri ⁴, Hesham Eissa ⁴, Frederick Yeo ⁴, Wassim Chemaitilly ⁵, Matthew J Ehrhardt ⁶, Johnnie Bass ⁷, Michael W Bishop ⁴, Kyla Shelton ³, Lu Lu ³, Sujuan Huang ³, Zhenghong Li ³, Eric Caron ³, Jennifer Lanctot ³, Carrie Howell ³, Timothy Folse ⁶, Vijaya Joshi ⁸, Daniel M Green ³, Daniel A Mulrooney ⁶, Gregory T Armstrong ⁶, Kevin R Krull ⁹, Tara M Brinkman ⁹, Raja B Khan ¹⁰, Deo K Srivastava ¹¹, Melissa M Hudson ⁶, Yutaka Yasui ³, Leslie L Robison ³

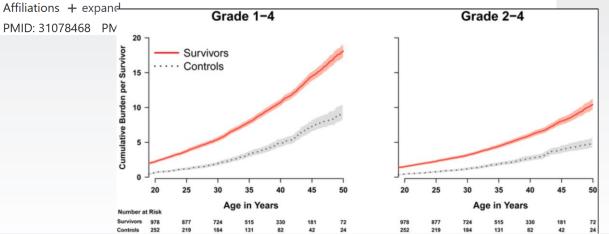
Affiliations + expand
PMID: 28890157 PMCID: PMC



> Lancet Haematol. 2019 Jun;6(6):e306-e316. doi: 10.1016/S2352-3026(19)30050-X. Epub 2019 May 8.

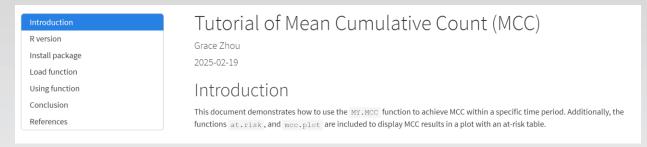
The changing burden of long-term health outcomes in survivors of childhood acute lymphoblastic leukaemia: a retrospective analysis of the St Jude Lifetime Cohort Study

Daniel A Mulrooney ¹, Geehong Hyun ², Kirsten K Ness ², Nickhill Bhakta ³, Ching-Hon Pui ⁴, Matthew J Ehrhardt ⁵, Kevin R Krull ⁶, Deborah B Crom ⁷, Wassim Chemaitilly ⁸, Deokumar K Srivastava ⁹, Mary V Relling ¹⁰, Sima Jeha ⁴, Daniel M Green ⁵, Yutaka Yasui ², Leslie L Robison ², Melissa M Hudson ¹¹

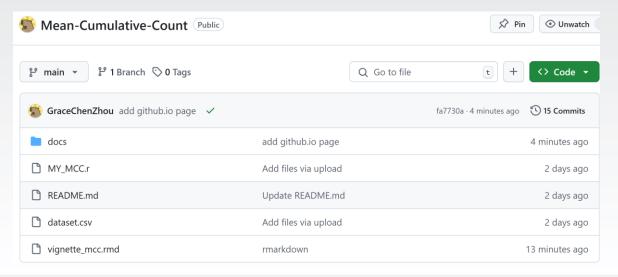




- Tutorial on MCC
 - Vignette: <u>Tutorial of Mean Cumulative Count (MCC)</u>



Reproducible source: https://github.com/GraceChenZhou/Mean-Cumulative-Count.git

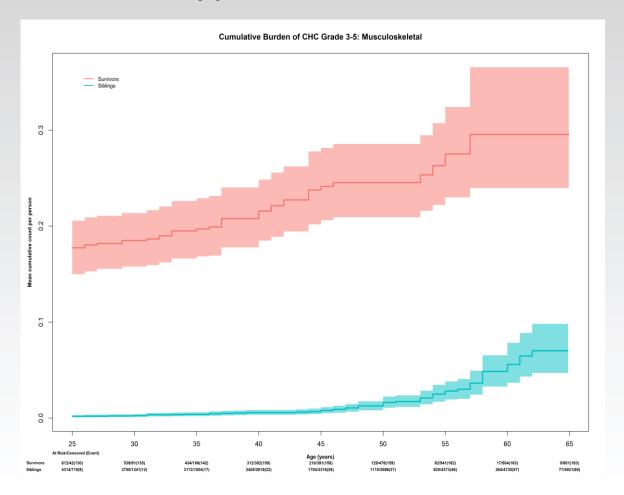




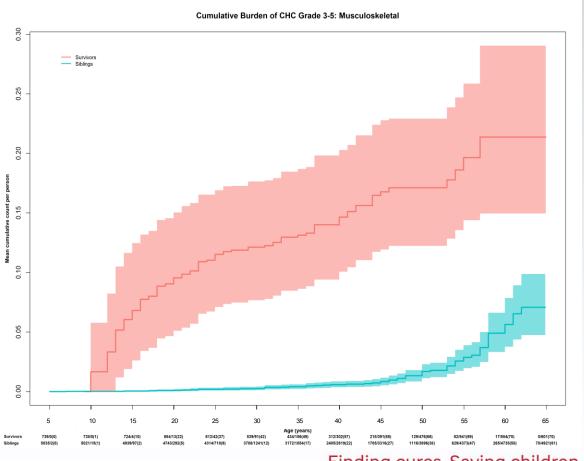
- Prevalent approach
 - Cutoff: 25 years old (diagnosis max age 20 + 5 years)
 - Aim: Eliminate the left-truncation
 - Caveat: Reduce the sample size
 - CIF: Subdistribution hazard model->effect on the CIF
 - HR: Subdistribution hazard model->effect on the subdistribution hazard function
- Weight approach (Geskus 2011)
 - Aim: Account for left-truncation
 - o CIF: Cause-specific cumulative incidence
 - HR: Proportional Fine-Gray subdistribution hazard model



Prevalent approach



Weight approach



Finding cures. Saving children.



Application: A CCSS study

- Subdistribution hazard ratio (95% CI)
 - Time to first Grade 3-5 MSK

Covariate	Prevalent approach (N=4929)	Weight approach (N=5776)
Group: Survivors (ref=Siblings)	21.63 (15.85, 29.52)	10.43 (7.03, 15.48)
Sex: Male (ref=Female)	1.14 (0.86, 1.5)	1.23 (0.87, 1.76)
Baseline age	1 (0.98, 1.03)	1.05 (1.03, 1.08)
Race/ethnicity: Hispanic (ref=NH White)	0.91 (0.45, 1.85)	1.03 (0.42, 2.52)
Race/ethnicity: NH Black (ref=NH White)	0.79 (0.19, 3.18)	1.03 (0.25, 4.2)
Race/ethnicity: Other (ref=NH White)	0.91 (0.48, 1.71)	0.62 (0.23, 1.69)



Conclusion

- Weight method for handling left-truncated data with competing risk in survival analysis
- Practical function for MCC estimation and visualization
- Correct mstate::crprep to obtain accurate weight (crprep2.r)
 - Usage: The same as the original <u>crprep</u> function from mstate R package
 - Location: MY_MCC.r
- R Shiny/R package will be developed based on the need



Acknowledgement

- Our team
 - Kumar, Kendrick, Sadie, Mengqi, Shalini, Mingjuan, Lu, Nivya, Zhuo
- Department of Epidemiology
 - Yan Chen
 - Geehong Hyun
 - Qi Liu



Reference

- Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Stat Med. 2017 Nov 30;36(27):4391-4400. doi: 10.1002/sim.7501. Epub 2017 Sep 15. PMID: 28913837; PMCID: PMC5698744.;
- Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks.
 Circulation. 2016 Feb 9;133(6):601-9. doi: 10.1161/CIRCULATIONAHA.115.017719. PMID: 26858290; PMCID: PMC4741409.
- Dong H, Robison LL, Leisenring WM, Martin LJ, Armstrong GT, Yasui Y. Estimating the burden of recurrent events in the presence of competing risks: the method of mean cumulative count. Am J Epidemiol. 2015 Apr 1;181(7):532-40. doi: 10.1093/aje/kwu289. Epub 2015 Feb 17. PMID: 25693770; PMCID: PMC4371763.
- Geskus RB. Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. Biometrics. 2011 Mar;67(1):39-49. doi: 10.1111/j.1541-0420.2010.01420.x. PMID: 20377575.
- Liesbeth C. de Wreede, Marta Fiocco, Hein Putter (2011). mstate: An R Package for the Analysis of Competing Risks and Multi-State Models. Journal of Statistical Software, 38(7), 1-30. URL https://www.jstatsoft.org/v38/i07/.
- R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical. Computing, Vienna, Austria. https://www.R-project.org/.



Questions



Appendix

•mstate::crprep

id	Tstart	Tstop	status	weight.cens	weight.trunc	count	failcode
1000218	20.59178	56.29863	0	1	1	1	1
1000228	24.84973	56	1	1	1	1	1
1000568	23.76164	27.44932	0	1	1	1	1
1000808	20.00546	49.02192	0	1	1	1	1
1001808	18.34973	38	1	1	1	1	1
1002038	15.66849	44.75956	0	1	1	1	1
1002408	16.88251	19.03836	2	1	1	1	1
1002408	19.03836	20	2	0.598027972	1	2	1
1002408	20	20	2	0.702820193	1	3	1
1002408	20	21	2	1.446924447	1	4	1
1002408	21	21	2	0.848254045	1	5	1
1002408	21	22	2	1.89192556	1	6	1
1002408	22	23	2	1.766662196	1	7	1
1002408	23	24	2	1.743029042	1	8	1
1002408	24	25	2	0.739810729	1	9	1
1002408	25	26	2	1.786706337	1	10	1
1002408	26	27	2	0.271863708	1	11	1
1002408	27	28	2	1.217781863	0.266934897	12	1
1002408	28	29	2	1	1	13	1
1002408	29	30	2	0.106638666	1	14	1
1002408	30	31	2	1.786706337	0.888311377	15	1
1002408	31	32	2	0.221929557	1	16	1
1002408	32	33	2	1.075247916	0.574279422	17	1
1002408	33	34	2	1.743029042	0.739947756	18	1
1002408	34	35	2	1.217781863	1	19	1
1002408	35	36	2	0.931330472	1	20	1
1002408	36	37	2	1.58503441	0.645785355	21	1
1002/08	27	38	2	1 63007197	1	າາ	1

crprep2

id	Tstart	Tstop	status	strata	weight.cens	weight.trunc	
1,000,218	20.00000	56.29863	0	1	1.0000000	1.000000]
1,000,228	24.00000	56.00000	1	1	1.0000000	1.000000	
1,000,568	23.00000	27.44932	0	1	1.0000000	1.000000	
1,000,808	20.00000	49.02192	0	1	1.0000000	1.000000	
1,001,808	18.00000	38.00000	1	1	1.0000000	1.000000	
1,002,038	15.00000	44.75956	0	1	1.0000000	1.000000	
1,002,408	16.00000	19.03836	2	1	1.0000000	1.000000	
1,002,408	19.03836	20.00000	2	1	0.9977376	1.000000	
1,002,408	20.00000	20.00000	2	1	0.9977376	1.141860	
1,002,408	20.00000	21.00000	2	1	0.9895592	1.141860	
1,002,408	21.00000	21.00000	2	1	0.9895592	1.241047	
1,002,408	21.00000	22.00000	2	1	0.9856633	1.241047	
1,002,408	22.00000	23.00000	2	1	0.9747619	1.353190	
1,002,408	23.00000	24.00000	2	1	0.9604919	1.422916	
1,002,408	24.00000	25.00000	2	1	0.9481546	1.479616	
1,002,408	25.00000	26.00000	2	1	0.9360652	1.541384	
1,002,408	26.00000	27.00000	2	1	0.9255640	1.541384	
1,002,408	27.00000	28.00000	2	1	0.9131824	1.541384	
1,002,408	28.00000	29.00000	2	1	0.8916708	1.541384	
1,002,408	29.00000	30.00000	2	1	0.8753077	1.541384	
1,002,408	30.00000	31.00000	2	1	0.8587187	1.541384	
1,002,408	31.00000	32.00000	2	1	0.8304082	1.541384	
1,002,408	32.00000	33.00000	2	1	0.7916112	1.541384	
1,002,408	33.00000	34.00000	2	1	0.7580516	1.541384	
1,002,408	34.00000	35.00000	2	1	0.7202485	1.541384	



Appendix

RE: Concerns about the crprep.default funciton



Ronald Geskus < rgeskus@oucru.org >

To: Zhou, Grace

Cc:
Srivastava, Deokumar

Start reply with: Sounds good, thank you. Sounds good, thanks! Thank you. I look forward to hearing from you.

Caution: External Sender. Do not open unless you know the content is safe.

Dear Grace,

Thanks for your email with the clear explanations. I'll have a closer look at it as soon as possible. Most likely after next week.

Best,

Ronald

