

# Appendix 1: Relationships between hazard rates for cancer incidence and subsequent death, with induced hazard rates for death following cancer occurrence

September 15, 2016

Denote by  $T_1$  the time from enrollment to cancer occurrence (i.e., diagnosis of invasive breast cancer or colorectal cancer) and by  $T_2$  the time from diagnosis to death (i.e., disease-specific death or death from any cause) among cancer patients. Let  $z = (z_1, \dots, z_p)$ , for  $p \geq 1$ , denote a vector of coded baseline participant characteristics. For example,  $z$  may include indicator variables for (all but one) body mass index (BMI) categories, along with variables that have potential to confound the relationship between BMI and the outcomes under study.

A Cox model hazard rate for  $T_1$  and time  $t_1$  following enrollment for a participant with baseline characteristics  $z$  can be written

$$\lambda_1(t_1; z) = \lambda_{10}(t_1) \exp\{z_1(t_1)\beta_1\}$$

where the modeled regression variable  $z_1(t_1) = \{z_{11}(t_1), z_{12}(t_1), \dots\}$  includes  $z$ , possibly along with product terms between elements of  $z$  and (data-analyst specified) functions of  $t_1$ , thereby allowing the hazard ratio for  $z$  to vary with  $t_1$ .

Similarly a Cox model hazard rate model for  $T_2$  at time  $t_2$  following cancer occurrence for a patient having cancer diagnosis at  $T_1 = t_1$  can be written

$$\lambda_2(t_2; z, t_1) = \lambda_{20}(t_2) \exp\{z_2(t_1, t_2)\beta_2\}$$

where the modeled regression variable  $z_2(t_1, t_2) = \{z_{21}(t_1, t_2), z_{22}(t_1, t_2), \dots\}$  could, for example, include  $z$ , indicator variables for categories of  $t_1$ , and product terms between elements of  $z$  and functions of  $(t_1, t_2)$ .

These models fully specify the hazard rate function for the composite outcome  $T_3 = T_1 + T_2$  given  $z$ , which represents time from enrollment to death (i.e., disease-specific or death from any cause), for individuals with baseline characteristics  $z$ . This induced hazard rate function is rather complex, however, but can be written at time

$T_3 = t_3$  as

$$\begin{aligned} \lambda_3(t_3; z) = & \int_0^{t_3} \lambda_{10}(t_1) e^{z_1(t_1)\beta_1} \exp \left\{ - \int_0^{t_1} \lambda_{10}(s_1) e^{z_1(s_1)\beta_1} ds_1 \right\} \\ & \lambda_{20}(t_3 - t_1) e^{z_2(t_1, t_3 - t_1)\beta_2} \exp \left\{ \int_0^{t_3 - t_1} \lambda_{20}(s_2) e^{z_2(t_1, s_2)\beta_2} ds_2 \right\} dt_1 \\ & \left[ \int_0^{t_3} \lambda_{10}(t_1) e^{z_1(t_1)\beta_1} \exp \left\{ - \int_0^{t_1} \lambda_{10}(s_1) e^{z_1(s_1)\beta_1} ds_1 \right\} \right. \\ & \left. \exp \left\{ \int_0^{t_3 - t_1} \lambda_{20}(s_2) e^{z_2(t_1, s_2)\beta_2} ds_2 \right\} dt_1 + \exp \left\{ - \int_0^{t_3} \lambda_{10}(s_1) e^{z_1(s_1)\beta_1} ds_1 \right\} \right]^{-1}. \end{aligned}$$

It is evident from this expression that proportional hazards models for  $T_1$  and  $T_2$ , is, for example, the special case given by  $z_1(t_1) = z$  and  $z_2(t_1, t_2) = z$ , generally do not imply a proportional hazards model for  $T_3$ . However, there are some useful special cases connecting regression models for  $T_1$  and  $T_2$  with that for  $T_3$ .

Specifically, we see that if  $\beta_1 = 0$  and  $\beta_2 = 0$  then  $\lambda_3(t_3, z) = \lambda_3(t_3)$  for all  $t_3 > 0$ , independent of  $z$ , so that null hypothesis for  $T_1$  and  $T_2$  imply a null hypothesis for  $T_3$ , in regard to association with the baseline regression vector  $z$ . Secondly if  $T_1$  is rare, so that  $\lambda_1(t_1; z)$  is close to zero for  $t_1$  in the study follow-up period, for all  $z$ , and therefore  $\exp\{-\int_0^{t_1} \lambda_1(s_1; z) ds_1\}$  is close to one for all  $(t_1; z)$ ; and  $\beta_2 = 0$ , so that the death rate for patients having the cancer under study doesn't depend on  $(t_1, z)$ , then the hazard rate  $\lambda_3(t_3; z)$  is approximately of proportional hazards form with regression coefficient  $\beta_1$ . This scenario applies approximately to the analyses of the paper. As a third special case suppose that both  $T_1$  and  $T_2$  are rare ( $\lambda_{10}$  and  $\lambda_{20}$  close to zero). One can then note from the above expression that  $\lambda_3(t_3; z)$  is of approximate proportional hazards form with regression coefficient  $\beta_3 = \beta_1 + \beta_2$ , assuming  $z_1(t_1) \equiv z$  and  $z_2(t_1, t_2) \equiv z$ . This scenario may also apply approximately to the analyses of this paper, though the disease-specific and all cause death rates among cancer patients are not so small during the follow-up of WHI cohorts.

There are a number of variants of the above models that may be useful in some applications. For example, dependencies of hazard rates on potential confounding variables may be modeled by stratifying baseline functions  $\lambda_{10}$  and  $\lambda_{20}$ , rather than by regression modeling, as is also the case for the dependence of  $\lambda_2$  and  $t_1$ . Also, another class of models would model the baseline hazard rate among cancer patients as a function of time since study enrollment, rather than time since cancer diagnosis, thereby giving a somewhat different interpretation to  $\beta_2$ , and to the induced hazard rate model for  $T_3$ . There is a considerable statistical literature on modeling choices for multivariate failure time variables of these types, and on related parameter estimation procedures (e.g., Prentice, Williams and Peterson, 1981; Andersen et al., 1993; Kalbfleisch and Prentice, 2002; Cook and Lawless, 2007).

#### Appendix References

Andersen PK, Borgan Ø, Gill RD, Keiding N. (1983). Statistical Models based on Counting Processes. New York: Springer Verlag.

Cook, RJ and Lawless JF. (2007). *The Statistical Analysis of Recurrent Events*. Springer.

Kalbfleisch JD and Prentice RL. *The Statistical Analysis of Failure Time Data*, Second Edition. New York: Wiley.

Prentice RL, Williams B, Peterson AV. (1981). On the regression analysis of multivariate failure time data. *Biometrika* 68, 373–379.

**eTable 1. Baseline Characteristics of Breast Cancer Survivor Cohort by baseline BMI group (n=7633)**

	Normal		Overweight		Grade-1 Obesity		Grade-2+3 Obesity		P <sup>1</sup>
	< 25		25 - < 30		30 - < 35		≥ 35		
	(n=2517)		(n=2658)		(n=1526)		(n=932)		
	N	%	N	%	N	%	N	%	
Age at screening									<0.001
50-59	837	33.3	796	29.9	446	29.2	319	34.2	
60-69	1110	44.1	1261	47.4	756	49.5	470	50.4	
70-79	570	22.6	601	22.6	324	21.2	143	15.3	
Race/ethnicity									<0.001
White	2299	91.3	2340	88.0	1299	85.1	741	79.5	
Black	86	3.4	156	5.9	141	9.2	136	14.6	
Hispanic	34	1.4	68	2.6	40	2.6	34	3.6	
American Indian	4	0.2	5	0.2	7	0.5	5	0.5	
Asian/Pacific Islander	75	3.0	59	2.2	20	1.3	7	0.8	
Unknown	19	0.8	30	1.1	19	1.2	9	1.0	
Education									<0.001
≤ High school/GED or less	356	14.2	483	18.3	341	22.5	216	23.4	
School after high school	784	31.4	998	37.8	627	41.3	412	44.5	
College degree or higher	1359	54.4	1159	43.9	550	36.2	297	32.1	
Hysterectomy at randomization	820	32.6	979	36.9	616	40.4	410	44.0	<0.001
Number of term pregnancies									<0.001
Never been pregnant/ No term pregnancy	363	14.6	377	14.3	178	11.8	118	12.7	
1	219	8.8	217	8.2	154	10.2	71	7.7	
2	750	30.1	676	25.6	338	22.3	198	21.3	
3	634	25.5	626	23.7	358	23.7	210	22.6	
4+	525	21.1	749	28.3	485	32.1	331	35.7	
Age at first birth									<0.001
Never pregnant/No term pregnancy	363	15.5	377	15.5	178	12.9	118	14.0	
<20	193	8.2	251	10.3	210	15.2	177	21.0	
20 – 29	1551	66.1	1569	64.7	863	62.4	466	55.3	
30+	240	10.2	229	9.4	133	9.6	81	9.6	
Family history of female relative with breast cancer	560	23.3	628	24.7	349	24.3	199	22.6	0.59
Bilateral oophorectomy	398	16.0	429	16.5	281	18.8	192	21.1	0.66
Treated diabetes (pills or shots)	34	1.4	69	2.6	85	5.6	92	9.9	<0.001
Smoking status									<0.001
Never	1190	48.0	1247	47.5	762	50.6	476	51.3	
Past	1098	44.3	1206	45.9	670	44.5	401	43.3	
Current	193	7.8	172	6.6	73	4.9	50	5.4	
Self-reported health									<0.001
Excellent	709	28.4	457	17.3	173	11.4	53	5.7	
Very good	1139	45.6	1238	46.9	614	40.4	294	31.6	
Good	559	22.4	784	29.7	583	38.4	418	44.9	
Fair/poor	90	3.6	161	6.1	149	9.8	165	17.7	
Duration of unopposed estrogen use									<0.001
None	1612	64.1	1724	64.9	1040	68.2	652	70.0	
Past User	300	11.9	312	11.7	178	11.7	112	12.0	
Current User	603	24.0	622	23.4	307	20.1	168	18.0	
< 5 Years ( <i>Duration; corresponds to past or current use</i> )	311	12.4	293	11.0	186	12.2	132	14.2	<0.001
5 - <10 Years	182	7.2	179	6.7	85	5.6	49	5.3	
10+ Years	420	16.7	462	17.4	215	14.1	99	10.6	
Duration of estrogen + progesterone use									<0.001
None	1455	57.8	1775	66.9	1140	74.7	725	77.8	
Past User	214	8.5	249	9.4	123	8.1	71	7.6	
Current User	847	33.7	631	23.8	263	17.2	136	14.6	
< 5 Years ( <i>Duration; corresponds to past or current use</i> )	443	17.6	363	13.7	193	12.6	110	11.8	<0.001
5 - <10 Years	303	12.0	270	10.2	104	6.8	59	6.3	
10+ Years	316	12.6	250	9.4	89	5.8	38	4.1	
CT participant	823	32.7	1184	44.5	829	54.3	552	59.2	<0.001
HT randomization group									0.85
CEE active	30	1.2	44	1.7	54	3.5	46	4.9	
CEE placebo	32	1.3	71	2.7	59	3.9	58	6.2	

<sup>1</sup> P-value is adjusted for age, race/ethnicity, education, and hysterectomy.

Cespedes Feliciano, et al. Methodological Considerations for Disentangling a Risk Factor's Influence on Cancer Incidence versus Post-Diagnosis Survival

CEE + MPA active	96	3.8	162	6.1	110	7.2	68	7.3	
CEE + MPA placebo	80	3.2	107	4.0	82	5.4	58	6.2	
DM randomization group									0.74
Intervention	239	9.5	369	13.9	243	15.9	158	17.0	
Comparison group	387	15.4	551	20.7	395	25.9	255	27.4	
	<b>Mean</b>	<b>(SD)</b>	<b>Mean</b>	<b>(SD)</b>	<b>Mean</b>	<b>(SD)</b>	<b>Mean</b>	<b>(SD)</b>	
Total energy expenditure/wk from physical activity (MET-hrs)	15.4	14.0	12.5	12.7	9.8	11.6	7.0	10.3	<0.001
RAND36 Physical functioning (0 -100 best)	88.1	14.3	83.9	17.0	77.5	20.4	67.8	23.4	<0.001
Height (cm)	163.3	6.3	162.4	6.2	161.8	6.2	161.1	6.7	<0.001
Weight (kg)	60.5	6.1	72.1	6.7	84.4	7.3	100.8	11.3	<0.001
Waist circumference (cm)	75.3	6.9	85.8	7.5	96.4	8.2	108.1	10.0	<0.001
Hip circumference (cm)	97.3	6.1	105.6	6.2	114.2	7.0	126.8	10.1	<0.001
Gail 5-year risk of breast cancer	2.0	1.1	2.0	1.2	2.0	1.2	1.8	1.2	0.58
Time to invasive breast cancer (years)	6.3	4.0	6.5	4.0	6.4	3.9	6.4	3.8	0.27

**eTable2. Baseline Characteristics of Colorectal Cancer Survivor Cohort by baseline BMI group (n=2290)**

	Normal		Overweight		Grade-1 Obesity		Grade-2+3 Obesity		P <sup>2</sup>
	< 25		25 - < 30		30 - < 35		≥ 35		
	(n=705)		(n=808)		(n=478)		(n=299)		
	N	%	N	%	N	%	N	%	
Age at screening									0.001
50-59	140	19.9	130	16.1	105	22.0	76	25.4	
60-69	315	44.7	384	47.5	229	47.9	161	53.8	
70-79	250	35.5	294	36.4	144	30.1	62	20.7	
Race/ethnicity									<0.001
White	630	89.4	679	84.0	391	81.8	226	75.6	
Black	26	3.7	71	8.8	63	13.2	60	20.1	
Hispanic	13	1.8	23	2.8	13	2.7	6	2.0	
American Indian	1	0.1	5	0.6	3	0.6	0	0.0	
Asian/Pacific Islander	23	3.3	17	2.1	5	1.0	4	1.3	
Unknown	12	1.7	13	1.6	3	0.6	3	1.0	
Education									<0.001
≤ High school/GED or less	139	19.9	184	22.9	130	27.3	76	25.6	
School after high school	270	38.6	330	41.1	185	38.8	125	42.1	
College degree or higher	290	41.5	289	36.0	162	34.0	96	32.3	
Hysterectomy at randomization	275	39.0	341	42.2	203	42.5	130	43.5	0.31
Number of term pregnancies									0.002
Never been pregnant/ No term pregnancy	103	14.7	94	11.7	49	10.4	27	9.1	
1	64	9.2	60	7.5	40	8.5	21	7.0	
2	173	24.7	175	21.8	96	20.3	60	20.1	
3	167	23.9	190	23.6	119	25.2	80	26.8	
4+	192	27.5	285	35.4	168	35.6	110	36.9	
Age at first birth									0.01
Never pregnant/No term pregnancy	103	16.2	94	12.8	49	11.1	27	10.0	
<20	57	8.9	102	13.9	76	17.2	50	18.5	
20 – 29	420	65.9	477	65.1	282	63.8	171	63.3	
30+	57	8.9	60	8.2	35	7.9	22	8.1	
Family history of female relative with breast cancer	135	20.2	133	17.3	74	16.8	57	20.1	0.57
Bilateral oophorectomy	126	18.3	149	18.9	98	21.3	64	22.1	0.28
Treated diabetes (pills or shots)	14	2.0	42	5.2	35	7.3	51	17.1	<0.001
Smoking status									0.003
Never	336	48.6	384	48.2	248	52.7	148	49.8	
Past	290	41.9	347	43.6	202	42.9	129	43.4	
Current	66	9.5	65	8.2	21	4.5	20	6.7	
Self-reported health									<0.001
Excellent	163	23.3	114	14.1	51	10.7	17	5.7	
Very good	303	43.3	362	44.9	186	39.1	86	28.9	
Good	190	27.1	272	33.7	189	39.7	140	47.0	
Fair/poor	44	6.3	58	7.2	50	10.5	55	18.5	
Duration of unopposed estrogen use									<0.001
None	464	65.9	501	62.0	339	70.9	227	75.9	
Past User	105	14.9	153	18.9	61	12.8	40	13.4	
Current User	135	19.2	154	19.1	78	16.3	32	10.7	
< 5 Years ( <i>Duration; corresponds to past or current use</i> )	99	14.0	122	15.1	63	13.2	44	14.7	<0.001
5 - <10 Years	50	7.1	57	7.1	23	4.8	11	3.7	
10+ Years	92	13.0	128	15.8	54	11.3	17	5.7	
Duration of estrogen + progesterone use									<0.001
None	536	76.2	641	79.3	392	82.2	266	89.0	
Past User	54	7.7	61	7.5	35	7.3	16	5.4	
Current User	113	16.1	106	13.1	50	10.5	17	5.7	
< 5 Years ( <i>Duration; corresponds to past or current use</i> )	80	11.3	89	11.0	41	8.6	20	6.7	<0.001
5 - <10 Years	38	5.4	41	5.1	26	5.4	6	2.0	
10+ Years	51	7.2	37	4.6	19	4.0	7	2.3	
CT participant	285	40.4	395	48.9	256	53.6	163	54.5	<0.001
HT randomization group									0.80
CEE active	27	3.8	33	4.1	20	4.2	21	7.0	
CEE placebo	17	2.4	32	4.0	21	4.4	18	6.0	

<sup>2</sup> P-value is adjusted for age, race/ethnicity, education, and hysterectomy.

Cespedes Feliciano, et al. Methodological Considerations for Disentangling a Risk Factor's Influence on Cancer Incidence versus Post-Diagnosis Survival

CEE + MPA active	37	5.2	48	5.9	32	6.7	13	4.3	
CEE + MPA placebo	47	6.7	53	6.6	24	5.0	27	9.0	
DM randomization group									0.71
Intervention	70	9.9	127	15.7	75	15.7	45	15.1	
Comparison group	111	15.7	158	19.6	121	25.3	71	23.7	
	<b>Mean</b>	<b>(SD)</b>	<b>Mean</b>	<b>(SD)</b>	<b>Mean</b>	<b>(SD)</b>	<b>Mean</b>	<b>(SD)</b>	
Total energy expenditure/wk from physical activity (MET-hrs)	14.0	14.2	11.5	11.9	10.0	12.2	6.4	9.8	<0.001
RAND36 Physical functioning (0 -100 best)	85.8	16.3	81.2	18.3	75.4	21.4	69.1	22.6	<0.001
Height (cm)	162.7	6.2	161.7	6.0	161.8	6.3	161.2	6.5	<0.001
Weight (kg)	60.4	6.1	71.8	6.7	84.6	7.3	101.5	11.5	<0.001
Waist circumference (cm)	76.8	8.3	87.0	7.3	97.9	7.7	109.4	10.7	<0.001
Hip circumference (cm)	97.0	5.5	105.4	6.3	114.3	6.8	127.3	9.9	<0.001
Gail 5-year risk of breast cancer	1.9	1.1	1.9	1.1	1.8	1.0	1.7	1.1	0.91
Time to colorectal cancer (years)	6.6	4.0	6.4	3.9	6.5	3.9	6.4	3.9	0.79

**eTable3. Breast Cancer Incidence and Death after Breast Cancer in the Women's Health Initiative: Multivariable-Adjusted<sup>3</sup> Hazard Ratios by BMI at Enrollment compared to women of normal weight (BMI < 25)**

	Normal < 25		Overweight 25 - <30				Grade-1 Obesity 30 - <35				Grade-2+3 Obesity ≥ 35				P-value <sup>4</sup>
	N	%	N	%	HR	CI	N	%	HR	CI	N	%	HR	CI	
<b>A. Events for the Full cohort; Time from enrollment to:</b>															
Invasive Breast Cancer(T1)	2517	(0.41)	2658	(0.43)	1.12	(1.06,1.18)	1526	(0.47)	1.31	(1.23,1.40)	932	(0.50)	1.48	(1.36,1.61)	<0.001
Deaths due to Breast Cancer (T3a)	194	(0.03)	226	(0.03)	1.11	(0.91,1.36)	137	(0.04)	1.19	(0.94,1.52)	124	(0.06)	1.99	(1.54,2.56)	<0.001
Deaths due to any cause after Breast Cancer (T3b)	409	(0.06)	450	(0.06)	1.09	(0.95,1.26)	296	(0.08)	1.36	(1.16,1.60)	214	(0.10)	1.89	(1.57,2.28)	<0.001
<b>B. Events for the Survivor cohort; Time from diagnosis of invasive breast cancer (post-diagnosis survival) to:</b>															
Deaths due to Breast Cancer (T2a) <sup>5</sup>	175	(0.98)	204	(1.13)	1.05	(0.84,1.30)	125	(1.20)	0.95	(0.73,1.23)	106	(1.73)	1.25	(0.94,1.67)	0.33
Deaths due to any cause (T2b)	409	(2.29)	450	(2.49)	0.99	(0.85,1.15)	296	(2.85)	0.99	(0.83,1.19)	214	(3.49)	1.25	(1.02,1.53)	0.10

<sup>3</sup> Covariate adjustments described in Figure 2 legend.

<sup>4</sup> Corresponds to a 1 degree-of-freedom test for trend of the association between BMI group and invasive breast cancer, death or survival.

<sup>5</sup> The difference in the number of events for deaths due to breast cancer after enrollment (i.e., 194, 226, 137 and 124) and deaths due to breast cancer after diagnosis (i.e., 175, 204, 125 and 106) is due to the participants that did not consent to extended follow-up but death information was available through passive follow-up sources. For these participants, time to death after enrollment is known, but time to incident invasive breast cancer was not available.



**eTable 4. Colorectal Cancer Incidence and Death after Colorectal Cancer in the Women's Health Initiative: Multivariable-Adjusted<sup>6</sup> Hazard Ratios by BMI at Enrollment compared to women of normal weight (BMI < 25)**

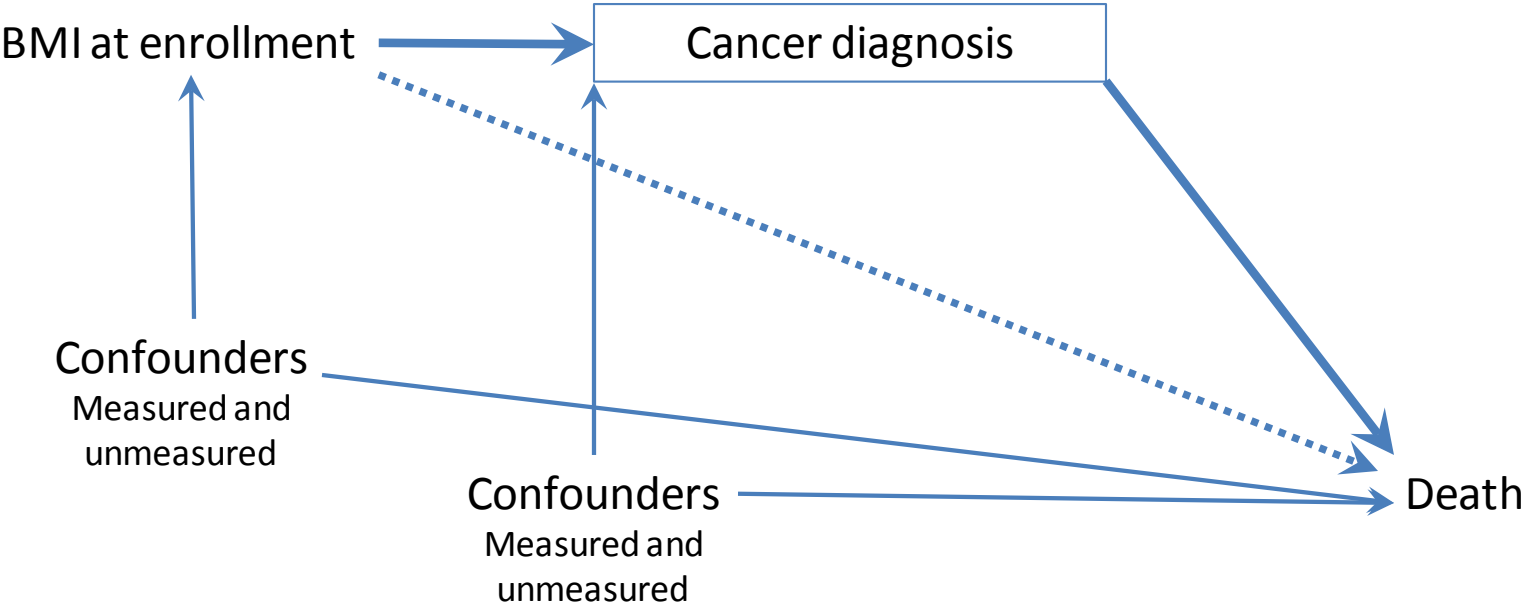
	Normal < 25		Overweight 25 - <30				Grade-1 Obesity 30 - <35				Grade-2+3 Obesity ≥ 35				P-Value <sup>7</sup>
	N	%	N	%	HR	CI	N	%	HR	CI	N	%	HR	CI	
<b>A. Events for the Full cohort;</b>															
<b>Time from enrollment to:</b>															
<b>enrollment (cancer incidence and cancer mortality)</b>															
Colorectal Cancer (T1)	705	(0.11)	808	(0.13)	1.10	(0.98, 1.23)	478	(0.14)	1.11	(0.96, 1.27)	299	(0.16)	1.20	(1.02, 1.42)	0.03
Deaths due to Colorectal Cancer (T3a)	202	(0.03)	238	(0.03)	1.18	(0.96, 1.46)	143	(0.04)	1.13	(0.87, 1.46)	100	(0.05)	1.40	(1.04, 1.88)	0.05
Deaths due to any cause after Colorectal Cancer (T3b)	245	(0.04)	306	(0.04)	1.17	(0.97, 1.42)	183	(0.05)	1.21	(0.96, 1.52)	120	(0.05)	1.41	(1.08, 1.85)	0.01
<b>B. Events for the Survivor cohort;</b>															
<b>Time from diagnosis of colorectal cancer (post-diagnosis survival) to:</b>															
Deaths due to Colorectal Cancer (T2a) <sup>8</sup>	182	(4.82)	212	(4.84)	1.09	(0.87, 1.38)	128	(4.95)	1.06	(0.79, 1.41)	82	(5.00)	1.18	(0.84, 1.66)	0.39
Deaths due to any cause (T2b)	245	(6.49)	306	(6.99)	1.15	(0.94, 1.41)	183	(7.08)	1.15	(0.90, 1.47)	120	(7.32)	1.27	(0.95, 1.69)	0.11

<sup>6</sup> Covariate adjustments described in Figure 3 legend.

<sup>7</sup> Corresponds to a 1 degree-of-freedom test for trend of the association between BMI group and colorectal cancer, death or survival.

<sup>8</sup> The difference in the number of events for deaths due to colorectal cancer (i.e., 202, 238, 143 and 100) and survival from colorectal cancer (i.e., 182, 212, 128 and 82) is due to the participants that did not consent to extended follow-up but death information was available through passive follow-up sources. For these participants, time from enrollment to death is known, but time to incident colorectal cancer was not available.

eFigure 1.





eFigure 3.

