

Non-Parametric Estimation of Reference Adjusted, Standardised Probabilities of All-Cause Death and Death Due to Cancer for Population Group Comparisons

Mark J Rutherford (✉ mark.rutherford@le.ac.uk)

University of Leicester

Therese M-L Andersson

Karolinska Institutet

Tor Åge Myklebust

Møre and Romsdal Hospital Trust

Bjørn Møller

Cancer Registry of Norway

Paul C Lambert

University of Leicester

Research Article

Keywords: age-standardization, net survival, crude probability of death, competing risks

Posted Date: April 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-429687/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Medical Research Methodology on January 6th, 2022. See the published version at <https://doi.org/10.1186/s12874-021-01465-w>.

**Title: Non-parametric estimation of reference adjusted,
standardised probabilities of all-cause death and death due to
cancer for population group comparisons**

Authors: Mark J Rutherford^{1,*}, Therese M-L Andersson², Tor Åge Myklebust^{3,4}, Bjørn Møller³, Paul C Lambert^{1,2}.

¹Department of Health Sciences, University of Leicester, UK.

²Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden.

*³ Department of Registration, Cancer Registry of Norway, Institute of Population-Based
Cancer Research, Oslo, Norway.*

*⁴ Department of Research and Innovation, Møre and Romsdal Hospital Trust, Ålesund,
Norway.*

* - corresponding author.

E-mail: mark.rutherford@le.ac.uk

Phone: +44 (0)116 2297292

1 **Abstract**

2 **Background:** Ensuring fair comparisons of cancer survival statistics across population groups
3 requires careful consideration of differential competing mortality due to other causes, and adjusting
4 for imbalances over groups in other prognostic covariates (e.g. age). This has typically been achieved
5 using comparisons of age-standardised net survival, with age standardisation addressing covariate
6 imbalance, and the net estimates removing differences in competing mortality from other causes.
7 However, these estimates lack ease of interpretability. In this paper, we motivate an alternative non-
8 parametric approach that uses a common rate of other cause mortality across groups to give
9 reference-adjusted all-cause and cancer-specific estimates rather than reporting net estimates.

10 **Methods:** We develop the methodology for a non-parametric equivalent of standardised and
11 reference adjusted crude probabilities of death, building on the estimation of non-parametric crude
12 probabilities of death. We illustrate the approach using regional comparisons of survival following a
13 diagnosis of rectal cancer for men in England. We standardise to the covariate distribution and other
14 cause mortality of England as a whole to offer comparability, but with close approximation to the
15 observed all-cause region-specific mortality.

16 **Results:** The approach gives comparable estimates to observed crude probabilities of death, but
17 allows direct comparison across population groups with different covariate profiles and competing
18 mortality patterns. In our illustrative example, we show that regional variations in survival following
19 a diagnosis of rectal cancer persist even after accounting for the variation in deprivation, age at
20 diagnosis and other cause mortality.

21 **Conclusions:** The methodological approach of using standardised and reference adjusted metrics
22 offers an appealing approach for future cancer survival comparison studies and routinely published
23 cancer statistics. Our non-parametric estimation approach through the use of weighting offers the
24 ability to estimate comparable survival estimates without the need for statistical modelling.

25 **Keywords:** age-standardization, net survival, crude probability of death, competing risks.

1 **Background**

2 Net survival measures are typically used for population-based cancer data as they enable fair
3 comparisons across population groups which have differential competing risks due to deaths from
4 causes other than cancer[1,2]. Past comparisons of individuals diagnosed with a specific cancer have
5 been made between groups defined by geographical areas[3-6], calendar time[7] or by population
6 subgroupings; such as age[8], socioeconomic status[9] or race[10]. In the context of cancer survival,
7 net survival measures the survival in the hypothetical world where it is not possible to die from
8 causes other than the cancer of interest. However, net survival measures have been criticized as
9 lacking a directly relevant interpretation, with some cautioning against relying on metrics that do not
10 “stick to this world”[11]. A further consideration when comparing population subgroups is that care
11 should be taken to ensure that the observed covariate distribution (age is the one primarily
12 considered) in each group is similar, or are enforced through some form of weighting or
13 standardization. See, for example, Corazziari *et al.*[12] for accounting for age distribution
14 differences. The same approach can be applied for other key covariates, dependent on the cancer
15 site and question of interest.

16
17 Crude probability measures (also referred to as cause-specific cumulative incidence functions in the
18 competing risks literature) offer a more interpretable metric, and have the advantage of being a
19 real-world measure. The cancer-specific crude probability of death measures the risk of dying of a
20 cancer at a particular timepoint in the presence of the competing risks due to other causes of death.
21 However, crude probabilities are a function of both the cancer-specific (or excess) mortality rate and
22 the other cause mortality rate, and so comparison between population groups are not “fair” when
23 trying to isolate differences solely due to the impact of cancer. This lack of “fairness” motivates the
24 use of net survival in the first place, rather than relying on the all-cause survival across groups, which
25 is also impacted by both competing mortality rates. Crude probability metrics have received some

1 recent attention in the relative survival framework, with various estimation approaches
2 proposed[13-16] and their use in a number of applied contexts[17-22].
3
4 Lambert *et al.*[23] propose estimation of all-cause survival and crude probability measures where
5 differences between population groups only depend on differences in excess (cancer) mortality
6 rates; they use the terminology of reference adjusted measures. In order to estimate the reference-
7 adjusted crude probability and all-cause measures in a relative survival (excess mortality)
8 framework, Lambert *et al.*[23] propose using common other-cause mortality rates from a reference
9 population when converting back from the excess mortality scale (which may well be the other-
10 cause mortality rates from one of the groups of interest). This leads to metrics that may not reflect
11 the actual experience of a population group, but are comparable and offer improved interpretability.
12 With careful selection of the appropriate reference adjustment and standardisation, these metrics
13 can also closely reflect the real-world experience of the study population. It is important to note that
14 the relevant other cause mortality rate for each group are first used in the estimation of the excess
15 mortality rates, prior to using the common other cause mortality rates for all groups when
16 converting to the all-cause scale.

17

18 In this paper, we build on the ideas of Lambert *et al.*[23], through development of non-parametric
19 methods to estimate the same underlying estimands. The approach of Lambert *et al.* relies on a fully
20 parametric setting; requiring the correct specification of the functional form for non-linear and time-
21 dependent covariate effects in a modelling framework. Our proposed non-parametric alternative
22 removes the requirement of the correct model specification. We apply the developed non-
23 parametric estimators and explore regional differences in survival following a diagnosis of rectal
24 cancer in England. This approach also builds upon previous research from Cronin and Feuer[24] for
25 estimating crude probability metrics, but we instead apply reference population other-cause
26 mortality rates to ensure that the crude probability metrics are fair in terms of differential other-

1 cause mortality when comparing across population groups. In terms of implementation, we discuss
2 the calculation of the metrics in both continuous time (that is, at unique event times), and with
3 follow-up time grouped into intervals.

4 **Methods**

5 ***Statistical methods***

6 We develop our estimation approaches within a relative survival (excess mortality) framework. This
7 is the most common approach to survival estimation in population-based cancer data, largely
8 because of the unavailability or unreliability of dichotomising cause of death information into a
9 death either due to cancer or due to other causes, which would be required in a cause-specific
10 survival estimation framework. We start by considering the all-cause mortality rate, $h_i(t)$, for an
11 individual i at time from diagnosis, t , which is assumed can be partitioned into component parts;
12 that due to the background mortality rate, $h_i^*(t)$, typically defined by information from population
13 mortality files and the excess mortality rate, $\lambda_i(t)$, for mortality associated with the diagnosis of
14 cancer:

$$15 \quad h_i(t) = h_i^*(t) + \lambda_i(t).$$

16 The subscript i denotes that this partitioning will vary by individual patient characteristics that could
17 impact the background mortality, the cancer-specific mortality or both; such as age-at-diagnosis or
18 sex. On the survival scale, this is formulated as:

$$19 \quad S_i(t) = S_i^*(t)R_i(t).$$

20 It is common to report marginal measures in population subgroups or the population as a whole.
21 Much of the recent literature in this research area relates to calculating the appropriate weighting
22 required to arrive at the correct marginal non-parametric estimates for the various quantities of
23 interest.

24

25

26

1 **Marginal measures**

2 Pohar Perme *et al.*[2] detail the appropriate weighting that gives an unbiased estimate of the
3 marginal relative survival, and the so-called Pohar Perme estimator is now widely used in practice.
4 Rather than reporting the marginal relative survival, others have recommended the use of crude
5 probabilities; i.e. the partitioning of the all-cause probability of death ($F(t) = 1 - S(t)$) into the
6 probability of death due to cancer ($F_C(t)$) and due to other causes ($F_O(t)$):

7
$$F(t) = 1 - S(t) = F_C(t) + F_O(t)$$

8 Following a similar notation to Sasieni and Brentnall[25], let $N_i(t)$ be a counting process that starts
9 at 0 and jumps to 1 at the time when individual i dies, and $Y_i(t)$ be an at risk process – an indicator
10 of whether an individual is still at risk at time t (1 if so, 0 otherwise), effectively $Y_i(t) = I(T_i \geq t)$
11 with $I()$ an indicator function. We can define $dN_i(t) = Y_i(t)I(T_i = t)$, which counts the events
12 specifically at time, t . We can then sum over all individuals at time t ; let $dN(t) = \sum_{i=1}^n dN_i(t)$ be the
13 sum over all individual events at time t , and $Y(t) = \sum_{i=1}^n Y_i(t)$ be the total number of individuals at
14 risk at time t . Taking a continuous time approach in the non-parametric context; for all individuals i
15 ($i=1\dots N$) at risk at time t , we can define the marginal all-cause cumulative hazard, $\hat{H}(t)$, as:

$$\hat{H}(t) = \int_0^t \frac{dN(s)}{Y(s)} = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s)} \quad (1)$$

16 This can be used in the estimation of the observed all-cause probability of death, $F(t)$, above;
17 $F(t) = 1 - S(t) = 1 - \exp(-\hat{H}(t))$. We can also derive the cumulative expected hazard for each
18 individual using the mortality rate information for the general population, using the mortality rates
19 based on an individual's, i , characteristics at time, t (i.e matched on attained age, attained calendar
20 year, and demographic characteristics; such as sex, deprivation group etc.). The cumulative excess
21 hazard for each individual, $H_i^*(t)$, is estimated up to time, t , from the population mortality rate
22 information, $h_i^*(t)$:

23
$$H_i^*(t) = \int_0^t h_i^*(s) ds,$$

1 with the expected (population) survival for each individual given by $S_i^*(t) = \exp(-H_i^*(t))$. There are
 2 a number of options for averaging these estimates to arrive at a marginal expected survival for the
 3 population and the appropriate approach depends on the context[2,15].

4

5 **Reference adjusted all-cause measures**

6 Our aim is to obtain reference adjusted all-cause survival so that differences are solely due to
 7 differences in cancer (excess) mortality. To do this a second group of expected mortality rates needs
 8 to be defined, so we can estimate what the all-cause survival would be in a population with the
 9 reference expected mortality rates. Lambert et al.[23] recommend the use of a second common set
 10 of population mortality rates for the purpose of adjusting to a reference population. With mortality
 11 rate information for an individual based on the rates in the second set of population mortality rates
 12 denoted with $h_i^{**}(t)$, one can arrive at the population survival under the reference standard
 13 population for an individual, i , $S_i^{**}(t) = \exp\left(-\int_0^t h_i^{**}(s) ds\right) = \exp(-H_i^{**}(t))$.

14

15 Adopting a second set of population mortality rates will influence the calculation of the all-cause
 16 cumulative hazard. Through the introduction of $h_i^{**}(t)$, the overall covariate distribution and
 17 hypothetical population at risk at time t may differ dependent on the relative difference between
 18 $H_i^{**}(t)$ and $H_i^*(t)$. To counteract this, weights can be introduced with the relative contribution
 19 depending on the ratio of the two expected survival estimates at time t for any given covariate
 20 pattern that influences expected mortality – here denoted for each individual i . Therefore, defining
 21 the weights at each timepoint, t , as $w_i(t) = \frac{S_i^{**}(t)}{S_i^*(t)}$, we can arrive at a reference-adjusted estimate of
 22 the marginal all-cause hazard, $H_R(t)$, at time t :

$$23 \quad \widehat{H}_R(t) = \int_0^t \frac{\sum_{i=1}^n w_i(s) Y_i(s) \{dN_i(s) - dH_i^*(s) + dH_i^{**}(s)\}}{\sum_{i=1}^n w_i(s) Y_i(s)} \quad (2)$$

24

1 The formula here adapts the marginal all-cause hazard on the population defined in Equation (1)
2 above in two ways. Firstly, the $\{-dH_i^*(s) + dH_i^{**}(s)\}$ term replaces the population hazard assumed
3 to be acting in the population with that of the reference standard. Secondly, the weights, $w_i(s)$, are
4 used to up or downweight individual events and risktime from those that are over or under-
5 represented in the reference population compared to the observed population at time, t , in exactly
6 the same way as described by Sasieni and Brentnall[25] for their relative survival index. It is worth
7 noting a number of features of the relation given in equation (2) under specific conditions. When
8 $H_i^{**}(t) = H_i^*(t)$ for all t and i this formula collapses to the Nelson Aalen estimator expressed in
9 equation (1); that is with the reference population mortality rates being equivalent to the population
10 expected mortality does not alter the all-cause hazard (nor the consequent calculations of crude
11 probability estimates). Removing the $dH_i^{**}(s)$ for all t , we arrive at the net cumulative hazard
12 estimator proposed by Sasieni and Brentnall, and this is therefore the all-cause hazard extension of
13 their estimator. Further adaptations of the formula above can also arrive at the Pohar Perme net
14 cumulative hazard (setting $S_i^{**}(t) = 1$), or the Ederer II estimator (setting $w_i(t) = 1$), as noted by
15 Sasieni and Brentnall[25].

16

17 **Reference adjusted crude probability measures**

18 The reference-adjusted all-cause hazard can then be de-composed to give the reference-adjusted
19 crude probabilities of death due to cancer and other causes using the same weighting, $w_i(t)$.

20 Defining the marginal net cumulative hazard using the reference standard as $\widehat{\Lambda}_R(t)$:

$$21 \quad \widehat{\Lambda}_R(t) = \int_0^t \frac{\sum_{i=1}^n w_i(s) Y_i(s) \{dN_i(s) - dH_i^*(s)\}}{\sum_{i=1}^n w_i(s) Y_i(s)}$$

22 which is the estimator proposed by Sasieni and Brentnall. And the marginal expected cumulative

23 hazard for the population hazard with the reference standard, $\widehat{H}_R^{**}(t)$:

$$\widehat{H}_R^{**}(t) = \int_0^t \frac{\sum_{i=1}^n w_i(s) Y_i(s) \{dH_i^{**}(s)\}}{\sum_{i=1}^n w_i(s) Y_i(s)}$$

The relevant crude probability of death due to cancer, $F_R^C(t)$, and other causes, $F_R^O(t)$, under the reference adjustment can then be given:

$$F_R^C(t) = \int_0^t S_R(u-) d\Lambda_R(u)$$

$$F_R^O(t) = \int_0^t S_R(u-) dH_R^*(u),$$

With $S_R(t)$ being the all-cause survival function at time, t , under the reference standard (i.e.

$$S_R(t) = \exp(-\widehat{H}_R(t)).$$

8

9 **Standardisation of covariates**

10 Furthermore, it may also be necessary to standardise to a specific covariate pattern (such as an age
11 profile) to allow direct comparability between groups or across different studies. This can be
12 achieved with a modification to the weights, $w_i(t)$, with a multiplication through by time-fixed
13 weights equivalent to those that have been used in age-standardisation traditionally[26,25,27].

14 Redefining the weights as:

$$w_i(t) = w_i^B \left(\frac{S_i^{**}(t)}{S_i^*(t)} \right)$$

16 where w_i^B is a time-fixed weight calculated at diagnosis and reflects the ratio of covariate pattern for
17 individual i in the external population relative to the observed population. For instance, these pre-
18 weights have been used for external age-standardisation using the International Cancer Standard
19 Survival weights. In that case, the weights, w_i^B , correspond to an individual's age (and the age-group
20 to which they belong), and are a relative comparison of the proportions in each age-group between
21 the current sample and the external standard. Again, the combination of these two sets of weights
22 are also proposed by Sasieni and Brentnall[25], but we convert their standardised relative survival
23 estimates to an all-cause and crude probability setting. The weights can be applied to all of the

1 reference-adjusted measure described above, so that any differences between groups are only due
2 to differences in excess mortality rates.

3 ***Software implementation***

4 In this paper, we use a continuous time implementation of the above approach, with the time-
5 dependent weights, $w_i(t)$, re-calculated at each unique event time. The software implementation is
6 via a user-written package in Stata; `stpp`. Further details on the implementation are given in the
7 Appendix. Cronin and Feuer[24] offer a lifetable approximation (calculation in interval periods, such
8 as months) to continuous time calculation for the observed crude probability estimates. It would be
9 also possible to make similar adjustments to the Cronin and Feuer lifetable approximation approach
10 by applying the reference population mortality information, with the appropriate adjustments for
11 the interval nature of the calculation as introduced by Cronin and Feuer.

12

13 ***Illustrative example***

14 We select two regions in England to make regional comparisons of survival following rectal cancer
15 for men diagnosed in the calendar period 2007-2012, with follow-up information available until the
16 end of 2013. For the calculation of relative survival, we use a population lifetable stratified by age,
17 deprivation, region, and sex. Given known discrepancies in deprivation-specific relative survival for
18 rectal cancer[28], it is important to quantify if region-specific survival differences remain that are not
19 due to differences in the proportion of men diagnosed in each deprivation group in a given region or
20 due to differences in the age distribution of men diagnosed with rectal cancer across regions. There
21 are also regional variations in other-cause mortality in England[29-31]. We therefore adopt an
22 approach of reference-adjusted and standardised survival comparison. The reference expected
23 mortality rates are for men in England as a whole in 2012. For our standardisation approach we
24 standardise to the age (in 5 broad age-groups; 15-44, 45-54, 55-64, 65-74, 75+) and deprivation
25 distribution of men diagnosed with rectal cancer for England as a whole for the study period of
26 interest. In using a common reference for population mortality rates, and a common covariate

1 distribution for age and deprivation, we provide all-cause and crude probability estimates that offer
 2 a fair comparison of the survival impact following a diagnosis of rectal cancer across the government
 3 office regions in England. We select two regions with different covariate profiles for illustration; the
 4 North East and South East regions. In adopting this approach, we remove the impact of regional
 5 variation in other-cause mortality on our estimates, and also any regional variation in the age and
 6 deprivation distribution of diagnoses – both of which would otherwise impact on the all-cause and
 7 cancer-specific cumulative probabilities of death.

8 Results

9 Table 1 describes the cohort of men diagnosed with rectal cancer between the years 2007 and 2012,
 10 and further summarises the age and deprivation distribution separately by the two government
 11 office regions. There are substantial differences in the distribution of deprivation across the two
 12 regions of England (31% in the most deprived group in North East, compared to 5% in the South
 13 West). This likely largely reflects regional variations in the deprivation distributions in the population
 14 in each region, rather than differential rectal cancer incidence by deprivation group across the
 15 regions. The age profile of the incidence of rectal cancer is relatively similar across the regions. In
 16 England as a whole, over 34% of cases of rectal cancer are in those over the age of 75, the oldest
 17 age-group we consider.

18 *Table 1: description of cohort for diagnoses of rectal cancer for men diagnosed between 2007 and 2012 in England, and the*
 19 *two comparison regions*

Region	Age-group N (row percentage)					Deprivation Group N (row percentage)					Total N
	15-44	45-54	55-64	65-74	75+	1- least deprived	2	3	4	5 – most deprived	
North East	56 (2.2%)	200 (8.0%)	579 (23.1%)	815 (32.5%)	857 (34.2%)	356 (14.2%)	346 (13.8%)	440 (17.6%)	594 (23.7%)	771 (30.8%)	2,507
South East	165 (2.6%)	499 (7.7%)	1,420 (22.0%)	2,059 (31.9%)	2,321 (35.9%)	2,124 (32.9%)	1,612 (24.9%)	1,432 (22.2%)	965 (14.9%)	331 (5.1%)	6464
England (TOTAL)	1,033 (2.5%)	3,181 (7.7%)	9,293 (22.6%)	13,563 (33.0%)	14,051 (34.2%)	8,479 (20.6%)	9,091 (22.1%)	8,825 (21.5%)	7,910 (19.2%)	6,816 (16.6%)	41,121

1 Table 2 shows the 5-year probability of all-cause death following a diagnosis of rectal cancer for men
2 in two regions of England. The 5-year crude probability of death due to cancer is also shown. Two
3 estimates are given for each of the above metrics; i) the observed value in each region; and ii) the
4 standardised and reference adjusted value. The observed values are quite similar across the two
5 compared regions; the all-cause 5-year probability of death is 51.1% in the North East region and
6 50.8% in the South East. However, these values are based on the covariate distributions and other-
7 cause mortality rates of each region separately. Applying the reference-adjustment (all-England
8 rates in 2012), and the covariate standardisation to England as a whole makes a marked difference
9 for these two regions. In the North East region the estimates decrease when standardising; this is
10 largely driven by the shift in the deprivation distribution (see Table 1). In contrast, the estimates for
11 the South East increase when using the reference adjustment and standardisation. This is again
12 largely driven by the shift in deprivation distribution, but also by altering to the all-England reference
13 rates for other cause mortality. Comparing the regions when the covariate distribution and other
14 cause mortality have been standardised shows that the North East region has a lower all-cause
15 probability of death than the South East (48.4% vs 52.6% respectively).

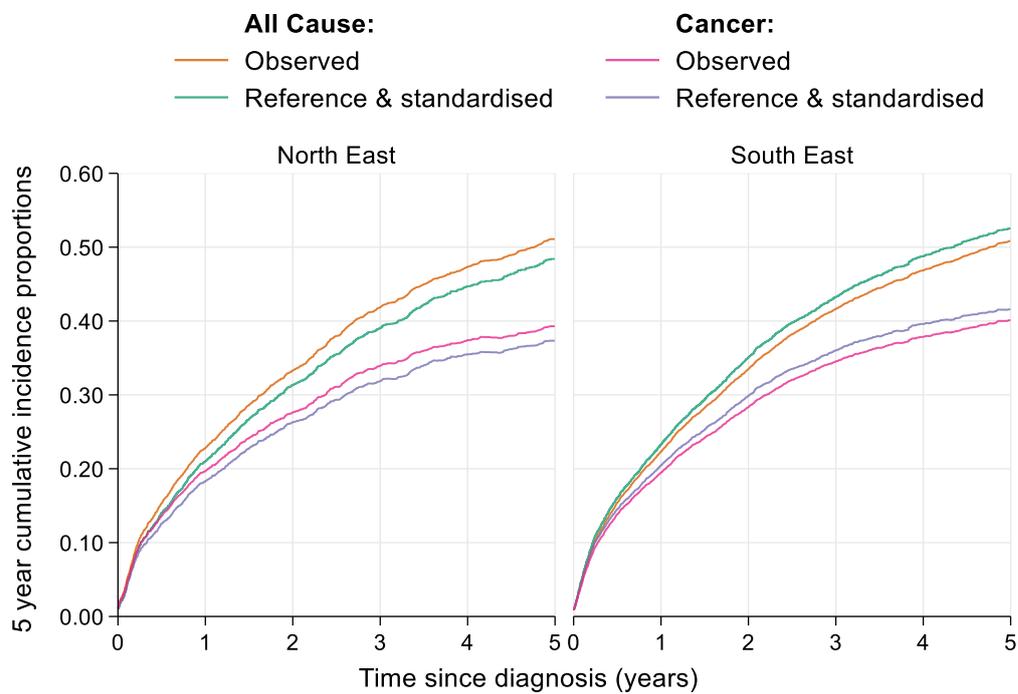
16

17 *Table 2: Regional values for the 5-year reference and standardised marginal probabilities of death (due to cancer and all-*
18 *causes) following a rectal cancer diagnosis for men (all ages). The reference standard is the population mortality rates for*
19 *men in England in 2012, and the standardisation group is the joint deprivation and age distribution of men with rectal*
20 *cancer in England as a whole. Also given for comparison, are values for the observed (unstandardized and non-reference*
21 *adjusted) probabilities.*

	Crude probability		All-cause death Probability	
	Due to cancer			
	Observed	Reference- adjusted and standardised	Observed	Reference- adjusted and standardised
North East	39.3%	37.3%	51.1%	48.4%
South East	40.1%	41.6%	50.8%	52.6%

22
23
24

1 Figure 1 shows the corresponding values across the entire range of time since diagnosis, as opposed
 2 to the point estimates given at 5-years in Table 2. The differential impact of reference adjustment
 3 for both the all-cause probability and death and the crude probability of death due to cancer can be
 4 clearly visualised. For the North East region, the reference adjustment results in a reduction in the
 5 probabilities of death; this is because a more favourable deprivation distribution is being applied
 6 through the standardisation. In contrast, for the South East, the reference and standardised
 7 estimates are higher across the entire range of follow-up than the corresponding observed values.



8

9 *Figure 1: Reference and standardised crude probability estimates (due to cancer and all-causes) for males diagnosed with*
 10 *rectal cancer in two English regions across time since diagnosis (all ages). The reference standard is the population*
 11 *mortality rates for men in England in 2012, and the standardisation group is the joint deprivation and age distribution of*
 12 *men with rectal cancer in England as a whole. Also given, are values for the observed (unstandardized and non-reference*
 13 *adjusted) probabilities.*

14

15

16

17

1 **Discussion**

2 Cancer survival metrics are typically calculated in the relative survival setting, and often externally
3 age-standardised marginal relative survival will be reported when comparing across regions or
4 countries (or stratifications for population subgroups) in order to remove the effect of differential
5 other-cause mortality. This measure is less relevant for patients and policy makers as it does not
6 reflect the real-world experience of cancer patients, however, it does allow for direct comparability
7 across groups with differential other cause mortality[32]. In this paper, we show that it is possible to
8 maintain this comparability whilst also retaining the overall burden of mortality (from deaths due to
9 cancer and other causes) when reporting metrics, but to do so requires the definition of a reference
10 standard for other cause mortality. The choice of the reference standard is important, but careful
11 choice means that the estimates reported can closely reflect the observed survival experience of the
12 cohort, whilst maintaining direct comparability.

13

14 Net survival metrics are popular in that, under assumptions of conditional exchangeability, these
15 completely remove the impact of other causes by effectively setting this to zero, and calculating the
16 survival function in the scenario where cancer is the only possible cause of death. This is often
17 achieved through estimation in the relative survival framework, where a further assumption of
18 relying on the population mortality rates to be the correct rates due to competing causes for the
19 cancer cohort. Our approach using this same estimation framework, but rather than assuming that
20 the rate due to competing causes is zero, we adopt a secondary reference population mortality file,
21 which is common across all comparison groups to allow fair real-world comparisons. This maintains
22 comparability, whilst improving the interpretability. Although, the exact interpretation of the
23 measure, is still a complex formulation, it often remains very close to the observed all-cause and
24 cancer-specific probabilities of death for each group of interest. In our example, we were able to
25 standardise to England as a whole for regional comparisons across England, which offers a logical of
26 choice common covariate distribution and other-cause mortality rate.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

A further important consideration when comparing across population groups is to ensure that a like-with-like comparison is undertaken for other key covariates that may differ in their distribution, but also impact on cancer survival. An obvious variable to consider is the age profile of the comparison groups, and it is common practice to age-standardise cancer survival estimates to account for this. The equivalent approach to standardisation is required for the reference adjusted metrics introduced in this paper. As with the reference standard for the other-cause mortality rates, a careful choice of covariate standard is needed to both allow direct comparability, whilst also providing survival estimates that are close to those observed in the population groups. Many previous research papers have adopted a common international age standard for cancer survival comparisons. The International Cancer Survival Standard weights are typically a younger profile than the cancer age distribution seen in England. On the relative survival scale, these difference in age profile often have a less dramatic difference, but on the all-cause scale, with age variation in both cancer and other-cause mortality, standardising to a an age profile that is younger can have a more marked difference. We further perform standardisation for the deprivation distribution in the regional comparisons in Table 2; it is often necessary to consider other key covariates to standardise over depending on the context of the comparison being made.

Although we standardise across the deprivation distribution to allow fair regional comparisons, this should not be seen as accepting the inequities in cancer survival seen across these population groups. This is done on the basis that to allow fair regional comparisons requires case-mix adjustment, which is only necessary due to the inequities in cancer (and other cause) survival across the deprivation groups. Another comparison of interest, that has been used in other studies, is to calculate the hypothetical gains in survival should the cancer-specific inequities in deprivation group survival could be removed. Furthermore, we have used regions in England that were formerly

1 referred to as Government Office Regions. These are fairly large population coverage areas (ranging
2 from ~2.5million to ~9 million individuals in 2019). For smaller geographical areas, there will be
3 much greater variation and a modelling approach that smooths through the survival estimates, and
4 also borrows strength across regions, would offer a better analysis strategy.

5

6 The non-parametric estimates can be calculated treating time continuously, or using a lifetable
7 approximation by splitting the timescale into pre-defined intervals (e.g. months). Applying the
8 grouped time approach reduces computational time in large datasets with little loss of accuracy if
9 the time intervals are sufficiently short. The fully continuous time approach requires the calculation
10 of the time-dependent weights at each event time, and so can become computationally intensive in
11 large datasets over long time periods. A further motivation for the lifetable approximation is when
12 this becomes necessary because of data availability reasons (e.g. survival information recorded to
13 the nearest month).

14

15 We have shown that non-parametric equivalents of standardised and reference adjusted survival
16 estimates can be obtained for cohorts of cancer patients. Using these metrics as a summary measure
17 of the burden of cancer and also for comparisons across groups is an approach that could be
18 adopted rather than reporting relative survival measures. The comparability ensured by using
19 relative survival (and age standardisation) is maintained, whilst also closely reflecting the observed
20 all cause survival patterns seen for the cohort of cancer patients; which can be broken down into the
21 relevant contributions of deaths due to cancer and other causes.

22

23

24

25

1 **Declarations:**

2 ***Ethics approval and consent to participate***

3 The study received ethical approval from North West - Greater Manchester East Research Ethics
4 Committee (14/NW/1449). The study was performed in accordance with the Declaration of Helsinki.

5 ***Consent for publication***

6 Not applicable.

7 ***Availability of data and materials***

8 The data that support the findings of this study are available from Public Health England
9 ([https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-](https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data)
10 [phe-odr-and-accessing-data](https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data)), but restrictions apply to the availability of these data, which were
11 used under license for the current study, and so are not publicly available.

12 ***Competing interests***

13 The authors declare that they have no competing interests.

14 ***Funding***

15 This work was supported by Cancer Research UK [Grant numbers: C41379/A27583 and
16 C1483/A18262]. The funding body had no role in the design of the study, analysis, and interpretation
17 of data or in writing the manuscript

18 ***Authors' contributions***

19 All authors contributed towards the conception and design of the study, and the interpretation of
20 data. PCL was further responsible for the acquisition of the data. MJR carried out the analysis of the
21 data and prepared the initial draft of the manuscript. All authors were involved in the substantial
22 revision of the manuscript and have approved the submission of the final manuscript.

23 ***Acknowledgements***

24 Not applicable.

1 **References**

- 2 [1] J. Estève, E. Benhamou, M. Croasdale, and L. Raymond, "Relative survival and the estimation
3 of net survival: Elements for further discussion," *Statistics in Medicine*, vol. 9, no. 5, pp. 529–538,
4 1990.
- 5 [2] M. P. Perme, J. Stare, and J. Estève, "On estimation in relative survival.," *Biometrics*, vol. 68,
6 pp. 113–120, Mar 2012.
- 7 [3] M. Coleman, D. Forman, H. Bryant, J. Butler, B. Rachet, C. Maringe, U. Nur, E. Tracey,
8 M. Coory, J. Hatcher, C. McGahan, D. Turner, L. Marrett, M. Gjerstorff, T. Johannesen, J. Adolfsson,
9 M. Lambe, G. Lawrence, D. Meechan, E. Morris, R. Middleton, J. Steward, and M. Richards, "Cancer
10 survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International
11 Cancer Benchmarking Partnership): An analysis of population-based cancer registry data," Jan. 2011.
- 12 [4] R. De Angelis, M. Sant, M. P. Coleman, S. Francisci, P. Baili, D. Pierannunzio, A. Trama,
13 O. Visser, H. Brenner, E. Ardanaz, M. Bielska-Lasota, G. Engholm, A. Nennecke, S. Siesling, F. Berrino,
14 R. Capocaccia, and EURO CARE-5 Working Group, "Cancer survival in europe 1999-2007 by country
15 and age: results of EURO CARE-5-a population-based study.," *Lancet Oncol*, vol. 15, pp. 23–34, Jan
16 2014.
- 17 [5] C. Allemani, T. Matsuda, V. Di Carlo, R. Harewood, M. Matz, M. Niksic, A. Bonaventure,
18 M. Valkov, C. J. Johnson, J. Estève, O. J. Ogunbiyi, G. Azevedo E Silva, W.-Q. Chen, S. Eser,
19 G. Engholm, C. A. Stiller, A. Monnereau, R. R. Woods, O. Visser, G. H. Lim, J. Aitken, H. K. Weir, M. P.
20 Coleman, and C. W. Group, "Global surveillance of trends in cancer survival 2000-14 (concord-3):
21 analysis of individual records for 37,513,025 patients diagnosed with one of 18 cancers from 322
22 population-based registries in 71 countries.," *Lancet (London, England)*, vol. 391, pp. 1023–1075,
23 Mar. 2018.
- 24 [6] M. Arnold, M. J. Rutherford, A. Bardot, J. Ferlay, T. M. Andersson, T. Å. Myklebust,
25 H. Tervonen, V. Thursfield, D. Ransom, L. Shack, *et al.*, "Progress in cancer survival, mortality, and

- 1 incidence in seven high-income countries 1995–2014 (icbp survmark-2): a population-based study,”
2 *The Lancet Oncology*, vol. 20, no. 11, pp. 1493–1505, 2019.
- 3 [7] M. Quaresma, M. P. Coleman, and B. Rachet, “40-year trends in an index of survival for all
4 cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-
5 2011: a population-based study.,” *Lancet*, vol. 385, pp. 1206–1218, Mar 2015.
- 6 [8] M. Araghi, M. Arnold, M. J. Rutherford, M. G. Guren, C. J. Cabasag, A. Bardot, J. Ferlay,
7 H. Tervonen, L. Shack, R. R. Woods, N. Saint-Jacques, P. De, C. McClure, G. Engholm, A. T. Gavin,
8 E. Morgan, P. M. Walsh, C. Jackson, G. Porter, B. M?ller, O. Bucher, M. Eden, D. L. O’Connell, F. Bray,
9 and I. Soerjomataram, “Colon and rectal cancer survival in seven high-income countries 2010-2014:
10 variation by age and stage at diagnosis (the icbp survmark-2 project).,” *Gut*, June 2020.
- 11 [9] A. Exarchakou, B. Rachet, A. Belot, C. Maringe, and M. P. Coleman, “Impact of national
12 cancer policies on cancer survival trends and socioeconomic inequalities in England, 1996-2013:
13 population based study,” *BMJ*, vol. 360, Mar. 2018.
- 14 [10] S. L. Stewart, R. Harewood, M. Matz, S. H. Rim, S. A. Sabatino, K. C. Ward, and H. K. Weir,
15 “Disparities in ovarian cancer survival in the united states (2001-2009): Findings from the concord-2
16 study,” *Cancer*, vol. 123, pp. 5138–5159, Dec. 2017.
- 17 [11] P. K. Andersen and N. Keiding, “Interpretability and importance of functionals in competing
18 risks and multistate models,” *Statistics in medicine*, vol. 31, no. 11-12, pp. 1074–1088, 2012.
- 19 [12] I. Corazziari, M. Quinn, and R. Capocaccia, “Standard cancer patient population for age
20 standardising survival ratios.,” *Eur J Cancer*, vol. 40, pp. 2307–2316, Oct 2004.
- 21 [13] P. C. Lambert, P. W. Dickman, C. P. Nelson, and P. Royston, “Estimating the crude probability
22 of death due to cancer and other causes using relative survival models,” *Statist. Med.*, vol. 29, no. 7-
23 8, pp. 885–895, 2010.
- 24 [14] M. P. Perme and K. Pavlic, “Nonparametric relative survival analysis with the r package
25 relsurv,” vol. 87.

- 1 [15] A. Belot, A. Ndiaye, M.-A. Luque-Fernandez, D.-K. Kipourou, C. Maringe, F. J. Rubio, and
2 B. Rachet, "Summarizing and communicating on survival data according to the audience: a tutorial
3 on different measures illustrated with population-based cancer registry data.," *Clinical epidemiology*,
4 vol. 11, pp. 53–65, 2019.
- 5 [16] D.-K. Kipourou, M. P. Perme, B. Rachet, and A. Belot, "Direct modeling of the crude
6 probability of cancer death and the number of life years lost due to cancer without the need of
7 cause of death: a pseudo-observation approach in the relative survival setting.," *Biostatistics*
8 (*Oxford, England*), May 2020.
- 9 [17] H. Charvat, N. Bossard, L. Daubisse, F. Binder, A. Belot, and L. Remontet, "Probabilities of
10 dying from cancer and other causes in French cancer patients based on an unbiased estimator of net
11 survival: a study of five common cancers.," *Cancer epidemiology*, vol. 37, pp. 857–863, Dec. 2013.
- 12 [18] B. K. Andreassen, T. . Myklebust, and E. S. Haug, "Crude mortality and loss of life expectancy
13 of patients diagnosed with urothelial carcinoma of the urinary bladder in Norway.," *Scandinavian*
14 *journal of urology*, vol. 51, pp. 38–43, Feb. 2017.
- 15 [19] S. I. Mozumder, P. W. Dickman, M. J. Rutherford, and P. C. Lambert, "InterPreT cancer
16 survival: A dynamic web interactive prediction cancer survival tool for health-care professionals and
17 cancer epidemiologists," *Cancer Epidemiology*, vol. 56, pp. 46–52, Oct. 2018.
- 18 [20] K. F. Wong, P. C. Lambert, S. I. Mozumder, J. Broggio, and M. J. Rutherford, "Conditional
19 crude probabilities of death for English cancer patients.," *British journal of cancer*, vol. 121, pp. 883–
20 889, Nov. 2019.
- 21 [21] P. Dasgupta, J. F. Aitken, C. Pyke, and P. D. Baade, "Competing mortality risks among women
22 aged 50-79 years when diagnosed with invasive breast cancer, Queensland, 1997-2012.," *Breast*
23 (*Edinburgh, Scotland*), vol. 41, pp. 113–119, Oct. 2018.
- 24 [22] P. Dasgupta, S. Cramb, K. Kou, X. Q. Yu, and P. D. Baade, "Temporal trends in net and crude
25 probability of death from cancer and other causes in the Australian population, 1984-2013.," *Cancer*
26 *epidemiology*, vol. 62, p. 101568, Oct. 2019.

- 1 [23] P. C. Lambert, T. M.-L. Andersson, M. J. Rutherford, T. Å. Myklebust, and B. Møller,
2 “Reference adjusted and standardized all-cause and crude probabilities as an alternative to net
3 survival in population-based cancer studies,” *International Journal of Epidemiology (in press)*, 2020.
- 4 [24] K. A. Cronin and E. J. Feuer, “Cumulative cause-specific mortality for cancer patients in the
5 presence of other causes: a crude analogue of relative survival,” *Statist. Med.*, vol. 19, no. 13,
6 pp. 1729–1740, 2000.
- 7 [25] P. Sasieni and A. R. Brentnall, “On standardized relative survival,” *Biometrics*, 2016.
- 8 [26] P. C. Lambert, P. W. Dickman, and M. J. Rutherford, “Comparison of different approaches to
9 estimating age standardized net survival,” *BMC Medical Research Methodology*, vol. 15, no. 1, pp. 1–
10 13, 2015.
- 11 [27] M. J. Rutherford, P. W. Dickman, E. Coviello, and P. C. Lambert, “Estimation of age-
12 standardized net survival, even when age-specific data are sparse,” *Cancer epidemiology*, vol. 67,
13 p. 101745, 2020.
- 14 [28] E. Syriopoulou, E. Morris, P. J. Finan, P. C. Lambert, and M. J. Rutherford, “Understanding the
15 impact of socioeconomic differences in colorectal cancer survival: potential gain in life-years,”
16 *British journal of cancer*, vol. 120, pp. 1052–1058, May 2019.
- 17 [29] L. M. Woods, B. Rachet, M. Riga, N. Stone, A. Shah, and M. P. Coleman, “Geographical
18 variation in life expectancy at birth in England and Wales is largely explained by deprivation.,”
19 *Journal of epidemiology and community health*, vol. 59, pp. 115–120, Feb. 2005.
- 20 [30] S. Walters, M. Quaresma, M. P. Coleman, E. Gordon, D. Forman, and B. Rachet,
21 “Geographical variation in cancer survival in England, 1991-2006: an analysis by cancer network.,” *J*
22 *Epidemiol Community Health*, vol. 65, pp. 1044–1052, Nov 2011.
- 23 [31] Office for National Statistics, “Cancer survival smoothed life tables. available at
24 [https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseas](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalsmoothedlifetables)
25 [es/datasets/cancersurvivalsmoothedlifetables,](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalsmoothedlifetables)” 2019.

1 [32] M. J. Rutherford, "Care needed in interpretation of cancer survival measures.," *Lancet*, Dec
2 2014.

3

4

Appendix

5

A1: Stata code for analysis

7 The estimates presented in this manuscript are implemented in the Stata user-written
8 package `stpp`, and are a continuous time representation of the estimates – with the
9 relevant weights and the contribution to the estimates calculated at each unique event
10 time.

11

12 Definition of selected variables:

13 `dx` - date of diagnosis

14 `dexit` - event date

15 `sex` - indicator variable for male/female

16 `yydx` - year of diagnosis

17 `status` - all-cause death indicator

18 `patid` - patient ID number

19 `extcomp` - variable with relative weights to the
20 age/deprivation group distribution in England as a
21 whole.

22

23 `popmort_region` - population mortality file stratified by age, region, sex and calendar
24 year.

25 `popmort_UK` - population mortality file stratified by age and sex only for England in
26 2012.

27

28 `stpp` is a user-written command. Type: `findit stpp` in Stata.

29

30 `*st-set` the survival data with 5 years follow-up**

31 `stset dexit, failure(status==1) exit(time`
32 `min(dx+`=5*365.24', mdy(12,31,2013))) origin(dx) id(patid)`
33 `scale(365.24)`

34

35 `**OBSERVED`

36 `stpp R_pp_o using popmort_region, ///`

37 `agediag(agediag) ///`

38 `datediag(dx) ///`

39 `by(region) ///`

40 `pmother(sex dep region) ///`

41 `allcause(AC_o) ///`

42 `crudeprob(CP_o OC_o) ///`

43 `deathprob list(5)`

44

```

1  **CREATES VARIABLES AC_O CP_O & OC_O - THE OBSERVED ALL-CAUSE,
2  CRUDE PROBABILITY OF DEATH DUE TO CANCER, AND CRUDE
3  PROBABILITY OF DEATH DUE TO OTHER CAUSES RESPECTIVELY.
4
5  **STANDARDISED & REFERENCE ADJUSTED
6  stpp R_pp_sr using popmort_region,          ///
7          ageddiag(ageddiag)                 ///
8          dateddiag(dx)                      ///
9          by(region)                          ///
10         pmother(sex dep region)             ///
11         using2(popmort_UK.dta,             ///
12              pmother2(sex)                 ///
13              pmyear2(.))                  ///
14         allcause(AC_sr)                    ///
15         crudeprob(CP_sr OC_sr)            ///
16         deathprob                          ///
17         indwei(extcomp) list(5)
18
19  **CREATES VARIABLES AC_sr CP_sr & OC_sr - THE STANDARDISED
20  (VIA EXTCOMP) & REFERENCE-ADJUSTED (VIA using2() OPTION) ALL-
21  CAUSE, CRUDE PROBABILITY OF DEATH DUE TO CANCER, AND CRUDE
22  PROBABILITY OF DEATH DUE TO OTHER CAUSES RESPECTIVELY.
23

```

Figures

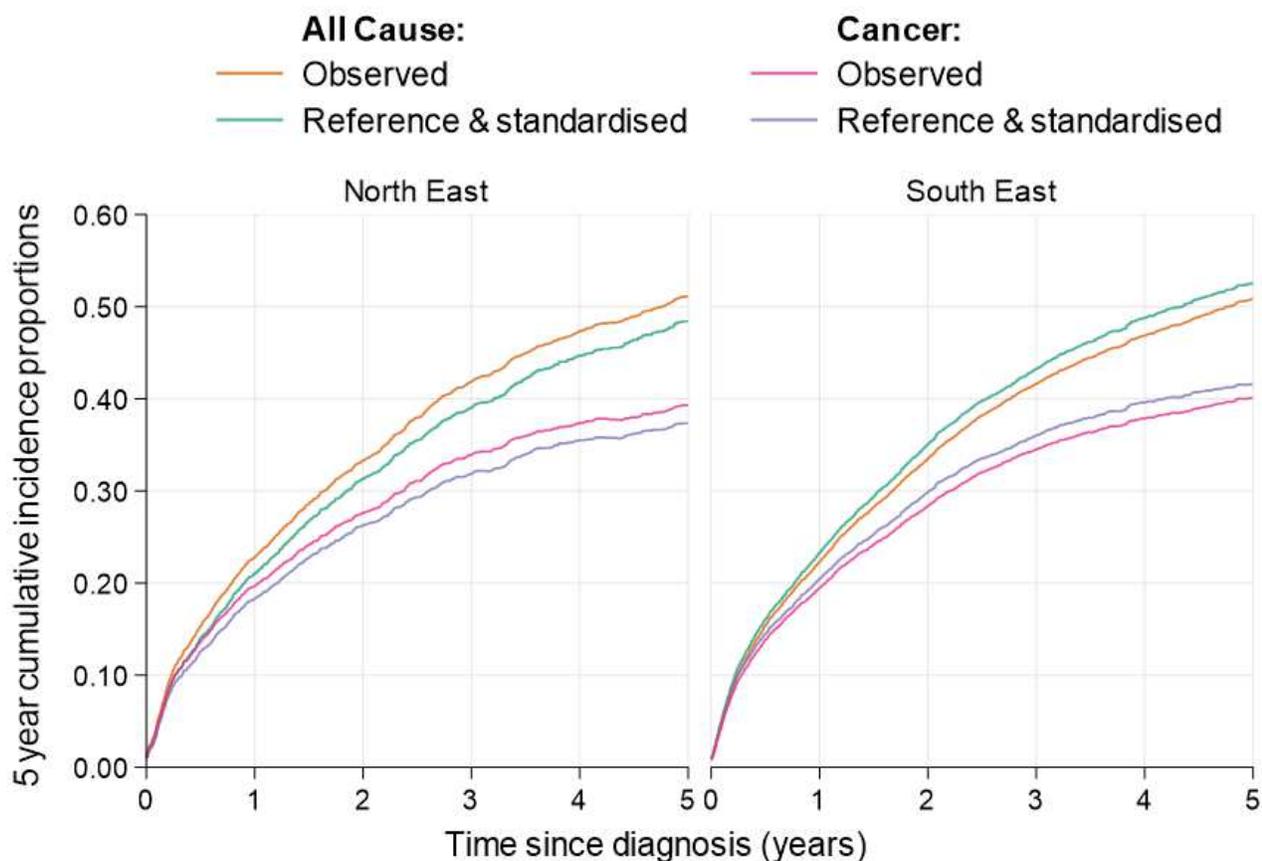


Figure 1

Reference and standardised crude probability estimates (due to cancer and all-causes) for males diagnosed with rectal cancer in two English regions across time since diagnosis (all ages). The reference standard is the population mortality rates for men in England in 2012, and the standardisation group is the joint deprivation and age distribution of men with rectal cancer in England as a whole. Also given, are values for the observed (unstandardized and non-reference adjusted) probabilities.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix.docx](#)