pHealth 2021
B. Blobel and M. Giacomini (Eds.)

doi:10.3233/SHTI210570

Communicating About Mortality in Health Decision Support: 'What and Why and When, and How and Where and Who'

Jack DOWIE[ab1], Mette Kjer KALTOFT[b], Vije Kumar RAJPUT[c]
[a]*London School of Hygiene and Tropical Medicine*
[b]*University of Southern Denmark*
[c]*Stonydelph Medical Centre, Tamworth, UK*

Abstract. The Covid-19 pandemic has only accelerated the need and desire to deal more openly with mortality, because the effect on survival is central to the comprehensive assessment of harms and benefits needed to meet a 'reasonable patient' legal standard. Taking the view that this requirement is best met through a multi-criterial decision support tool, we offer our preferred answers to the questions of What should be communicated about mortality in the tool, and How, given preferred answers to Who for, Who by, Why, When, and Where. Summary measures, including unrestricted Life Expectancy and Restricted Mean Survival Time are found to be reductionist and relative, and not as easy to understand and communicate as often asserted. Full lifetime absolute survival curves should be presented, even if they cannot be 'evidence-based' beyond trial follow-up limits, along with equivalent measures for other criteria in the (necessarily) multi-criterial decision. A decision support tool should relieve the reasonable person of the resulting calculation burden.

Keywords: mortality, life expectancy, survival curves, Restricted Mean Survival Time, Time-to-Event, multi-criteria decision, reasonable patient, preference-sensitive

1. Background

Our setting is the person making an individual health decision, often in conjunction with a healthcare professional, but often at home in the community (where one of their health decisions may be whether or not to contact a health professional). Our underlying motivation and aim is the development of a *generic* decision support *template*, to be the basis of decision support *tools* (DSTs) for *specific* health-related decisions. We take for granted that any health decision is multi-criterial, whether the objective is health promotion or disease prevention - often in the absence of a professional diagnosis - or therapeutic - often after a professional diagnosis. As eponymously indicated, the focus here is on the mortality criterion, a literally vital one, albeit always alongside other criteria such as those relating to morbidity and option burden. These DSTs must facilitate optimal *individualisation* and *personalisation*, in ways to be defined - and hence clearly distinguished.

[1] Corresponding author: Jack Dowie, LSHTM, 15-17 Tavistock Place, London, UK WC1H 9SH; email: jack.dowie@lshtm.ac.uk

This is a huge topic, so we use Kipling's simple taxonomy of 'What and Why and When, And How and Where and Who' to move the discussion along. It will immediately be apparent, however, that the possible answers to all of these are interrelated, indeed in many cases seriously interdependent. Our prime concern is with What and How, as the main determinants of the structure and contents of the decision support template. When, Where, Who, and Why address the background contexts and conditions. Throughout, the proffered answers reflect our *preferences*. Others may, and almost certainly will, have different preferences - and answers - reflecting beliefs and interests different from ours.

2. Who, When, Where

Who for? is the best place to start, since we need to determine and distinguish the parties involved in decision-supporting communication about mortality, and Who splits naturally into 'Who *for*?' and 'Who *by?*'. To us, the communication is *for* the *reasonable person* (RP) making a health decision. A subset of RP's health decisions involve those taken within professional healthcare services, following a decision to engage with them. RP is assumed to require no support in recognising that a health decision is required and acknowledges the mortality implications of their health-related options must be considered. They will therefore positively request and/or expect this within any decision support.

When is RP making a health decision? They are making a health *decision* whenever, after maximum possible *individualisation*, there is no option that has the best or equal best performance rates on all their decision criteria – 'the things that matter to them'. If there is such a 'dominant' option, there is actually no *decision* to make and so no need for *personalisation* to make it preference-sensitive. Individualisation is the process of refining the *option performance rates* for a criterion on the basis of the individual's characteristics. Personalisation is the process of adjusting the inputs to a decision on the basis of the individual person's preferences, i.e. their value judgements. Our preference is for these to be expressed as their *criterion importance weights*.

So, taking life expectancy (LE) as an example of a possible mortality criterion, in individualisation the estimated LE for an individual is progressively refined from an initially coarse (least individualised) one, based only on their sex and age, through adjustments in the light of other epidemiological characteristics (e.g. family history, ethnicity, location) and with reference to personally observable characteristics (e.g. weight, blood pressure), and, finally, to the most individualised one by adding those determinable mainly by healthcare professionals or services (biological markers, medication history). Personalisation may take two forms. In comparing the LE associated with two options, the importance attached to the extension offered by the superior option may not involve simple linear extrapolation: the importance of LE changes may be subjected to *time preference*, usually 'discounting'. Secondly, LE is assigned importance weight relative to other criteria in the decision. The degree of personalisation will reflect the extent to which the elicitation and incorporation of criterion and time preferences are facilitated in (the) decision support.

Who by? For us, the communication is *by* either a healthcare professional or a credible provider-independent source. This reflects our distinction between the RP-*as-person* and RP-*as-patient*, i.e. as 'reasonable patient', legally speaking [1]. That distinction leads us directly into that between *apomediative* and *intermediative* decision support [2, 3] - and so to our preferred answers to Where**.**

Where? Intermediative decision support usually comes in the form of a Patient Decision Aid (PDA), designed to help the clinician and patient decide - *in a clinic or teleconsultation* - what is best for the patient. The options included in a PDA are typically restricted to those which the clinician feels relevant to the patient, and those they are able to prescribe or recommend within a guideline. The delivery of the aid within the encounter is under the clinician's control. In contrast, an apomediative Decision Support Tool (DST) is a 'direct-to-person-*as-person'* resource designed to help the person decide what is best for themselves - *at home or elsewhere in the community,* and usually online. They are developed by a provider-independent team of professional health analysts, operating within consumer law and rights provisions rather than under health professional 'duty of care' commitments, legal and ethical. A clinician may become involved, subsequently, in adding their support to that provided by an apomediative DST that the person has consulted - and will add value to the extent that they have greater relevant knowledge and can enhance *individualisation*. However, the significance of their individualisation improvement will depend on *personalisation* - improving individualisation in relation to a lowly-weighted criterion may be of little importance. Finally, a clinician may actively promote 'hybrid apo-intermediation' by inviting patients to engage with an apomediative aid at home in advance of a clinic consultation. While this transforms the apomediative DST into a clinician-managed one, the resulting encounter will be different from pure intermediation, because of the different answers to What and How, and Why, to which we now turn.

3. Why, What, How

Why? The healthcare literature displays growing acceptance of the need to deal with mortality more directly, and openly, than has been traditional in clinical practice. The Covid-19 pandemic has only accelerated this process, because the effect on survival is central to the comprehensive assessment of option harms and benefits to which the person is legally entitled under a 'reasonable *patient'* standard. In the absence of an overriding reason, this standard makes clear that *some* communication about mortality is required, leading directly to the consequential questions of What and How regarding the contents of the 'some'. At the other extreme, simple consumer demand, again absent some overriding reason, provides sufficient grounds for *some* communication about mortality, either on request as intermediative patient, or on search as apomediative person. The law trumps, as in the case of *lack* of demand, or outright refusal of communication about mortality by the patient, the acquiescing clinician may not obtain informed consent and be at legal risk. [1]

What constitutes an 'overriding reason' becomes crucial. It will almost always take the form of an *ethical* argument of a consequentialist or deontological sort, perhaps endorsed or even mandated by a professional code. How particular ethical reasons play out within the 'reasonable patient' standard is likely to remain legally moot in our view. However, we reject any attempted ethical justification for the suppression or distortion of the best available mortality estimates achievable through *individualisation*. These include ones intended to maintain 'hope' and 'optimism', since a requisite understanding of probability provides a firmer and superior ethical foundation for the *warrantable* hope and optimism that RP is entitled to maintain. RP is not devoid of emotions, moods and feelings, but accepts that the decision support offered (and requested) should be minimally distorted in content and delivery by temporary - entirely 'natural' and 'human'

- affective states. At the same time, they are aware and assured their basic feelings constitute the necessary and valid basis for preference-sensitive decisions, as achieved through *personalisation*.

*What***?** We assume RP requires decision support incorporating the best (individualised) numerical central point estimates for a mortality measure, accompanied by credible intervals. RP accepts that introducing verbal quantifications – high/low, long/short, big/small - will not add value and increase the possibility of bias, cognitive or motivational.

Any mortality measure relates to a single option (alternative/strategy//intervention), and so, in the (genuine, undominated) decisional context, the preferred mortality measure is required *for each of these* options. This creates a key issue, unfortunately too often ignored or overlooked. Are the two (or more) option-specific mortality measures to be processed as a set, producing a *mono-criterial* comparative assessment of the relative mortality impact of the options (usually in the form of a *difference)*? Or are they to be kept in isolation from each other and processed separately, each in conjunction with the equivalent measures for the other criteria in the *multi-criterial* assessment that characterises most health decisions? The answer has serious implications for personalisation, as we now illustrate, in the course of introducing the most familiar example of one of the possible types of answer to What: Life Expectancy, a *summary mortality measure on an unrestricted time scale*.

Assume that the LE for a 70 year old man without an intervention is 81, and that with the intervention it is 84. A widespread practice in making such a binary decision is to *reduce* the two LEs to a single summary measure of mortality impact, by taking their difference. The effect of the intervention is seen as a LE increase of 3 years, maybe phrased as 3 Life Years Gained by the intervention, or Lost by not having it. However, if the person attaches importance to other criteria, such as avoiding morbidity and option burden, they need to compare their personalized overall assessment of a non-dominant intervention e.g. 'LE 84; morbidity 79; burden 65', with that for the no intervention scenario 'LE 81; morbidity 79; burden 75'. Premature reduction of absolute criterion values to their differences, rules out proper personalization through criterion importance weighting.

This fundamental issue remains latent as we go more deeply into the possible answers offered to What, starting with the term 'unrestricted'. Unrestricted means that the mortality measure is calculated over full lifetime, not for any limited or 'restricted' time horizon. LE is unrestricted in being calculated as the average length of life of all those in a cohort, including the years lived by those in the tail of its distribution, e.g. those surviving beyond 100. (Median LE is sometime used to reduce the effect of the more extreme outliers have on the Mean LE, but it is still an unrestricted measure.) While it may seem obvious that we require an unrestricted measure, some of the most used mortality measures are 'Restricted', including the increasingly popular Restricted Mean Survival Time (RMST) metric [4]. Its methodological home is in the wider 'Time-to-Event' (TTE) literature, which embraces events less final than death, especially adverse ones such as heart attack, stroke, or fracture. Many recent papers make the case for TTE metrics as being equal to, if not better than, conventional measures, such as the Hazard Ratio (HR), Relative Risk (RR), or Number Needed to Treat (NNT), in capturing and communicating the evidence-based merits of options to clinicians and patients. Most who advocate TTE measures do so because, in line with those conventional measures, they present trial results in a statistically valid way, even when follow-up falls well short of lifetime, as it usually does.

In briefly documenting the case being made for TTE measures when the event is death, we see that How is being addressed along with What. As explained by Kloecker:

"The RMST difference compares the areas under the 2 survival curves for the intervention and control groups for a specified (restricted) interval. This contrast corresponds to the mean temporal postponement of the outcome in one group compared with the other, with each group-specific RMST quantifying the average delay in the event over the specified time horizon… [Take as example a 70-year-old man.] On the basis of EMPA-REG OUTCOME [a trial with a 4 year follow-up], the Hazard Ratio for all-cause mortality is 0.68. The corresponding RMST difference is 21 days postponement over 4 years. The health care professional could advise the patient that empagliflozin, on average, would prolong his life by 21 days over 4 years" [5] (pp 541, 545).

There is internal debate as to whether RMST is best measured by the dominant 'vertical' method of Kloecker and others [6] or by the alternative 'horizontal' method first deployed by Lytsy in relation to the 4S statin trial [7] and used recently by Bellavia [8]. However, we do not need to enter this debate since, measured either way, restricted follow-up means that in most cases only a limited number of patients have experienced the target event by study closure. In the horizontal measure TTE can be calculated only for those percentiles of the population that have experienced the event (death) - a percentage that will often fall far short of even the median survival length (50th percentile). So, in the Lytsy statin study, the delay of 1 year calculated for the seventh percentile *by definition* applies only to the 7% in the untreated group who we know died within the follow-up period of the study. But we also know, again by definition, that this delay does not apply to the 93% who did not die.

Why use any *restricted* survival measure then? The answer reflects the preference for producing - and having guidelines and clinicians limited to using - 'evidence-based' mortality measures, where evidence is effectively restricted to that produced by the results of RCTs. Apart from the terminal situation, there is therefore only one possible exception: when the trial data permit 'unrestricted' conclusions. Claggett, et al. argue an 'actuarial' approach sometimes makes this possible. Used by them in two conditions [9,10], and by Dorresteijn [11], the 'delay of death' is measured from the person's *age at* diagnosis, not from *time of* diagnosis (irrespective of age). This method requires the underlying study/trial data to have "wide age range in the patient population, as well as a sufficiently large number of events occurring across the age spectrum in order to allow for relatively stable age-specific risk estimates… For any given age, a survival curve was estimated, representing the survival probabilities over time for patients alive at that age and receiving LCZ696 [the control]. A corresponding survival curve was then estimated using data from patients receiving enalapril [the intervention]. This process was repeated to produce estimated treatment effects for each age from 45-80 years old....[9](Supplementary Appendix, p2). The Utrecht group led by Dorresteijn applied this unrestricted lifetime TTE measurement to the results of the Women's Health study in which nearly 28,000 initially healthy women were followed up for a mean of 10 years. Crucially, there were a substantial number of older participants, because "predictions of lifetime models are not limited by the follow-up time of the study but rather by the age distribution of study participants. Therefore, observations in elderly patients are essential for stable long term predictions" [11] (p1).

So, an 'evidence-based' commitment means that the unrestricted lifetime mortality estimates to which RP is entitled can be provided only in the relatively rare cases meeting standard evidential requirements. However, we note that it is not only, or even primarily, the scientific standards for what constitutes 'evidence', that are the problem. Many of

the appropriate standards could be met, if it were not for the domination of short-term funding arrangements and commitments by the public and private supporters of research. For most researchers and groups, a circa 3-year 'restricted mean survival time' is the reality.

No particular restricted TTE is meaningful to a RP. Short of their lifetime, each and every particular restricted time period could be meaningful to a particular RP. There is therefore no way in which one of these can be privileged just because it is the follow-up limit of relevant robust, 'evidence-generating' research. RP is entitled to attach personal meaning to each and every extra year of life as represented in the full survival functions relevant to their decision. In practice and in life there is therefore no alternative to lifetime extrapolation beyond the restricted evidence. There are two possibilities. One is *explicit* extrapolation based on one of many possible principles, e.g. that the proportional hazard ratio calculated from the survival curves continues to apply. The other possibility is implicit, undiscussed, maybe unconscious, extrapolation. However, for the clinician, who, remember, is always engaging with RP, the latter is not a legally secure choice. They will need to supply, and document, their best explicit extrapolation, based on all their resources.

Despite emphasising its virtues, most advocates of RMST agree there is a serious issue with its simple use as an answer to What, because:

"… estimates of RMST differences depend on, and should be interpreted with reference to, the event rate in both groups and the duration of follow-up or, rather, the specified time horizon… Because the RMST difference reflects the difference in the areas under 2 survival curves, the same difference may be obtained from diverse combinations of survival curves. For example, an RMST difference of 1 year over 5 years may be the result of an RMST of either 4 years in the treatment group and 3 years in the control group or 3 years in the treatment group and 2 years in the control group. As with the aforementioned measures of absolute risk reduction, we therefore advise interpretation of the RMST along with the survival curves" [5] (p548).

This is too weak. The much stronger version, and our preferred answer to What, takes us immediately to the two (Kaplan-Meier) survival curves, from which all restricted or unrestricted single summary measures, such as RMST and LE, are derived. Referring to the underlying survival curves (restricted or unrestricted) cannot be merely an *ideal,* or something that *could* lead to *'better appreciation'* [12] (p58). All single summary mortality measures are completely dependent on the two generative survival curves, and the *absolute* probabilities (KM estimators), that compose those curves, are essential to give personal meaning to the difference between them. Any single number mortality metric such as the RMST difference is *reductionist* as well as *relative*; it removes the *two* absolute groundings required in a multi-criterial health decision - which means all health decisions involving benefits and harms. The plotting of survival probabilities and curves on a truncated scale, common in the TTE literature (e.g. from 0.95 to 1 in Kloecker), exacerbates this problem.

Let us *imagine* how Reasonable Person might respond to the RMST TTE measure, having had it explained in the context of the Lytsy statin study:

'So if - but only if - I were to die 6 years from now, after taking statins for that time, I would be dying one year later than I would have, if I hadn't taken statins for that period. I would have 'postponed my death' from 5 years to 6 years by taking statins - *if* I hadn't avoided dying. But it is very likely I will avoid dying in this period according to your chart of the two survival curves. It shows that at the end of the period shown - 6 years - my probabilities of survival are 87% without statins and 91% with them. This delay

calculation, which after considerable effort I do *understand*, might mean something if I knew I was fairly certain to die 6 years from now. However, I am certainly not certain to die in exactly 6 years - or any particular time from now. The difference between the two probabilities which generates the 1 year benefit from statins is 4%. But this difference has no real use for me either, because 7% and 11% would generate the same 1 year 'postponement' - but affect my decision very differently from 87% and 91%. My decision involves trading off any mortality benefit from statins with the absolute probabilities for my other benefits and harms, such as morbidity, adverse events, side effects, and treatment burdens from statins. In fact, it's now clear, no one pair of absolute probabilities on the survival curves (like 87% and 91% at 6 years) is of any use. What has unquestioned personal meaning is my maximum lifetime, so I need to keep in mind the full survival curves, not any summary measure and if some restricted time horizon has meaning for me, I will decide what it is!"

*How***?** The best available estimates of the KM survival curves for the individual, for each option, are our starting point for decision supporting any RP. Restricted, reductionist, relative measures may be introduced, but only subsequently, never initially, and never as substitutes in the belief (we believe unwarranted) that they are easier to understand and communicate than probability-based measures. Throughout the TTE literature it is suggested that summary measures based on a *time scale*, should replace continuous measures based on a *probability* scale.

"It is well known that the established effect measures are associated with some difficulties when used in clinical care for individual decision-making…Probabilistic thinking is difficult. Laymen, patients, and even skilled professionals all suffer from various degrees of statistical illiteracy, making it difficult for many to perform simple arithmetic calculations and to comprehend risk estimates. This predicament is further supported by research showing that the format of the effect measure may influence patients' acceptance of taking a medication as well as doctors' and health authorities' willingness to recommend or prescribe it. This signifies the challenge clinicians face when deciding how to describe treatment outcomes to their patients for the purpose of shared decision-making" [7] (p 905).

We argue the reverse: a time scale measure should be introduced, *if at all*, only as a supplement to the fundamental probabilistic formulation. Neither ease of understanding or communication (*if true*) can justify prioritising a non-probabilistic measure, derived from a probabilistic one, on the grounds that it reduces the difficulty of accepting and dealing with the inherent uncertainty (correctly) reflected in that underlying one. Succumbing to the tempting reductionist relative alternative is a cognitive version of the 'streetlight fallacy', where keys lost in the dark are searched for under the light, because it is easier to see there.

Perhaps surprisingly, many of the most explored issues in relation to How loom much less large for us than others. Our purpose in presenting the survival curves is not to have them interpreted, even if there is some evidence that they can be well understood [13, 14]. Indeed, we positively discourage attempts at interpretation of visual displays of all kinds, including emoticon displays that follow best visualization practice in providing information *on a single criterion*. This is because *calculation* on their basis is essential and multi-criterial calculation is not a task to be imposed on an individual engaging with a decision support tool, whether highly numerate or not. Implying they are capable of it amounts to symbolic violence [15]. Reasonable Person should know what the various data elements represent and what procedure ('algorithm') is being used to integrate them, but not have to attempt synthesizing calculation. The purpose in presenting survival

curves is to *help the RP understand them only to the extent necessary to understand why analytical calculation based on them - in conjunction with equivalent lifetime measures for other criteria - is essential*. And so to understand why the multi-criteria scores resulting from these calculations should be the core output of personalized decision support tools.

References

[1] UK Supreme Court Judgment, Montgomery v Lanarkshire Health Board (2015) [Available at https://www.supremecourt.uk/cases/docs/uksc-2013-0136-judgment.pdf
[2] Eysenbach G. From intermediation to disintermediation and apomediation: new models for consumers to access and assess the credibility of health information in the age of Web2.0. Stud Health Technol Inform. 2007; 129: 162–6.
[3] Dowie J, Kaltoft M.K. The future of health is self-production and co-creation based on apomediative decision support. Med Sci. 2018; 6: 66.
[4] Nemes S, Bülow E, Gustavsson A. A brief overview of Restricted Mean Survival Time estimators and associated variances. Stats 2020; 3: 107–19.
[5] Kloecker DE, Davies MJ, Khunti K, Zaccardi F. Uses and limitations of the restricted mean survival time: Illustrative examples from cardiovascular outcomes and mortality trials in type 2 diabetes. Ann Intern Med. 2020; 172: 541–52.
[6] Eloranta S, Smedby KE, Dickman PW, Andersson TM. Cancer Survival Statistics for Patients and Healthcare Professionals – a Tutorial of Real-World Data Analysis. J Int Med 2020; 289: 12-28.
[7] Lytsy P, Berglund L, Sundström J. A proposal for an additional clinical trial outcome measure assessing preventive effect as delay of events. Eur J Epidemiol. 2012; 27: 903–9.
[8] Bellavia A, Wallentin L, Orsini N, James SK, et al. Time-based measures of treatment effect: Reassessment of ticagrelor and clopidogrel from the PLATO trial. Open Heart 2017; 4: e000557
[9] Claggett B, Packer M, McMurray JJV, Swedberg K, et al. Estimating the long-term treatment benefits of sacubitril-valsartan. N Engl J Med. 2015; 373: 2289–90.
[10] Claggett B, Lachin JM, Hantel S, Fitchett D, et al. Long-term benefit of empagliflozin on life expectancy in patients with type 2 diabetes mellitus and established cardiovascular disease survival estimates from the Empa-Reg OUTCOME trial. Circulation. 2018; 138: 1599–1601.
[11] Dorresteijn JAN, Kaasenbrood L, Cook NR, van Kruijsdijk RCM, et al. How to translate clinical trial results into gain in healthy life expectancy for individual patients. BMJ. 2016; 352: 1–6.
[12] Belot A, Ndiaye A, Luque-Fernandez MA, Kipourou DK, et al. Summarizing and communicating on survival data according to the audience: a tutorial on different measures illustrated with population-based cancer registry data. Clin Epidemiol. 2019; 11: 53–65.
[13] Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Mortality versus survival graphs: Improving temporal consistency in perceptions of treatment effectiveness. Patient Educ. Couns. 2007; 66: 100–107.
[14] Rakow T, Wright RJ, Bull C, Spiegelhalter DJ. Simple and multistate survival curves: Can people learn to use them? Med Decis Mak. 2012; 32: 792–804.
[15] Kaltoft MK, Nielsen JB, Salkeld G, Dowie J. Health informatics can avoid committing Symbolic Violence by recognizing and supporting generic decision-making competencies. Stud Health Tech Inform. 2015; 218: 172-7.