Estimating the 95% confidence interval for survival gain between an experimental anti-cancer treatment and a control

Ther Adv Med Oncol
2017, Vol. 9(11) 721–723
DOI: 10.1177/1758834017731084

Andrea Messori, Erminia Caccese and Maria Claudia D'Avella

Correspondence to:
Dr Andrea Messori
HTA Unit, ESTAR Toscana, Regional Health System, via San Salvi 12, 50100 Florence, Italy
andrea.messori.it@gmail.com

Erminia Caccese
Maria Claudia D'Avella
HTA Unit, ESTAR Toscana, Regional Health System, Florence, Italy

A growing number of new anticancer treatments are being approved by the European Medicines Agency (EMA) and/or the US Food and Drug Administration (FDA), and for this reason oncologists and decision makers are increasingly requested to interpret appropriately the results of survival studies. The mean survival (or mean lifetime survival) is in theory the best parameter to represent survival outcome in a given group of patients,[1,2] but in most cases mean survival cannot be estimated from censored observations (unless the event rate reaches 100% in the patients concerned or unless the survival curve is extrapolated to infinity through complex mathematical functions such as those of Weibull[3] or Gompertz[4]). In this framework, median survival is almost universally employed in clinical studies focused on time-to-event endpoints (irrespective of whether the endpoint is overall survival [OS] or progression-free survival [PFS]). Accordingly, median survival is considered an adequate proxy for mean survival.

On the other hand, survival gains have long been considered the index that best represents the incremental survival benefit observed between two treatments aimed at the same disease condition.[5] In theory, in two-arm studies, the survival gain should be determined as the difference between the two values of mean survival (i.e. gain = mean $\text{survival}_{\text{experimental group}}$ – mean $\text{survival}_{\text{control group}}$), but in practice, the gain is determined as the difference between the two values of median survival (i.e. gain = median $\text{survival}_{\text{experimental group}}$ – median $\text{survival}_{\text{control group}}$). Again, this applies indistinctly to OS and PFS.

When it comes to determining the 95% confidence interval (CI) of these parameters, medians do not raise any computational problems: in fact, the most commonly used statistical programmes carry out this estimation of 95% CIs and, more importantly, the articles describing these clinical studies generally report this information. For example, in the recent approval document of olaratumab[6] (available from the EMA website), the pivotal trial comparing olaratumab *versus* placebo in 133 patients with soft tissue sarcoma presents the following information on PFS (see Figure 7 on page 65 of the European Public Assessment Report [EPAR] document[6]):

(1) experimental arm: number of patients = 66, median PFS = 6.6 months with 95% CI from 4.1 to 8.3 months;
(2) control arm: number of patients = 67, median PFS = 4.1 months with 95% CI from 2.8 to 5.4 months;
(3) hazard ratio (HR) of progression for experimental *versus* control arms = 0.67 with 95% CI from 0.44 to 1.02 (p = 0.0615).

Interestingly, the difference in PFS is at the limits (p = 0.0615) of the conventional threshold of statistical significance (p = 0.05).

The difference in PFS (i.e. the PFS gain) is clearly 2.5 months (i.e. 6.6–4.1 months), but the 95% CI for this difference is not reported. The 95% CI for the PFS gain is extremely useful for interpreting the result of this clinical study and, furthermore, this information is mandatory for incorporating the clinical studies into the GRADEpro software,[7] which is the web tool developed to permit the application of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) method. This latter implication is particularly relevant in Italy because the Italian Medicines Agency recognizes the characteristics of innovativeness only in agents that have been evaluated according to the GRADE method.[8]

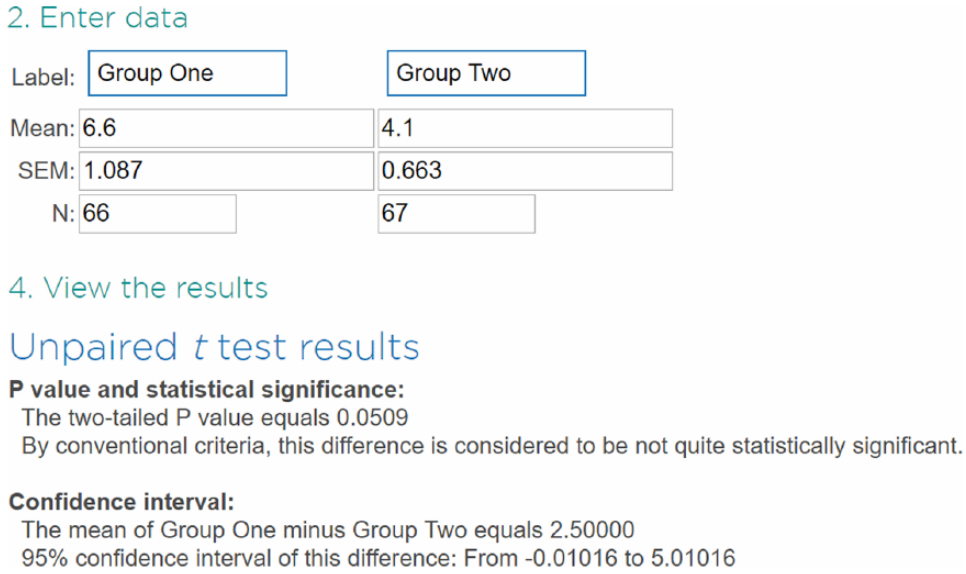

Figure 1. Screenshots from the online tool developed by GraphPad: both sections of 'Enter data' and 'View the results' show the survival data for our example on olaratumab. SEM, standard error of the mean.

One simple method of estimating the 95% CI of the gain in OS or PFS proceeds through the two following steps:

Step 1: the standard error of the mean (SEM) for survival length (or length of PFS) is separately calculated for the two patient arms according to the following well-known equation:

$$\mathrm{SEM} = \frac{\begin{pmatrix}\text{upper limit of 95\%}\\ \text{CI for median survival}\end{pmatrix} - \begin{pmatrix}\text{lower limit of 95\%}\\ \text{CI for median survival}\end{pmatrix}}{1.96 \times 2}$$

hence, SEM is 1.087 for the experimental group, where 1.087 = (8.3–4.1)/(1.96×2), and SEM is 0.663 for the control group, where 0.663 = (5.4–2.8)/(1.96×2).

Step 2: a widely used internet tool, GraphPad software,[9] is applied, that converts these data of survival length (namely: median PFS = 6.6 months, SEM = 1.087, and number of patients = 66 for the experimental arm; median PFS = 4.1 months, SEM = 0.663, and number of patients = 67 for the control arm) into an estimate of the gain in PFS (gain = 2.5 months), accompanied by its 95% CI (from −0.01016 to +5.01016 months). Figure 1 shows how this internet tool presents the input data and the results of this example.

It should be noted that there is a small discrepancy between the level of statistical significance originally reported in the olaratumab EPAR (p = 0.0615 associated with the HR analysis) and that resulting from the procedure described herein (p = 0.0509). This discrepancy is related to the assumption by which median survival is considered to be a proxy of mean survival.

Generally speaking, marked discrepancies between the results of traditional HR analyses and those based on median survival gains are unlikely. However, should these discrepancies occur, interpreting these discrepancies will lead to identifying some specific characteristics of the survival dataset that can explain the discrepancy itself; for example, the presence of a small subset of long-term survivors can explain why the HR provides a more favourable interpretation of the results than the median gain.

As metrics for managing survival data, median survival has several well-known disadvantages. For example: (a) it leads to loss of information, as it represents only one time point in the survival curve; (b) it is a poor indicator of treatment benefit when hazards are nonproportional, as may occur with targeted therapy; (c) it cannot be computed when 50% of residual survival has not been reached in at least one trial arm; (d) its standard error is quite large for commonly used sample sizes of a few hundred patients. On the other hand, the main disadvantage of the HR

approach is that the metrics cannot be used to carry out a cost-effectiveness analysis (CEA), because all CEAs require knowledge of the survival gain between the two treatments. Extensive literature has emphasized this advantage of gains in pharmacoeconomic terms.[5,10,11] Hence, using the gains implies a series of well-known disadvantages, but there is also an important advantage.

In conclusion, the procedure described in this report can be useful to reaffirm the important role of gains in survival studies and to improve the current situation in which most survival assessments are based exclusively on the values of HR. According to our proposal, the estimation of median survival gain (with 95% CIs) should not replace the traditional interpretation based on HR (with its 95% CI and/or *p* value), but should be seen as an additional piece of information that improves the overall interpretation of the survival data.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not for-profit sectors.

Conflict of interest statement

The authors declare that there is no conflict of interest.

References

1. Messori A, Trippoli S, Becagli P, *et al.* Pharmacoeconomic profile of paclitaxel as a first-line treatment for patients with advanced ovarian carcinoma. A lifetime cost-effectiveness analysis. *Cancer* 1996; 78: 2366–2373.
2. Messori A. Survival curve fitting using the Gompertz function: a methodology for conducting cost-effectiveness analyses on mortality data. *Comput Methods Programs Biomed* 1997; 52: 157–164.
3. Messori A, Becagli P and Trippoli S. Median versus mean lifetime survival in the analysis of survival data. *Haematologica* 1997; 82: 730.
4. Hoyle MW and Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. *BMC Med Res Methodol* 2011; 11: 139–152.
5. Wright JC and Weinstein MC. Gains in life expectancy from medical interventions – standardizing data on outcomes. *N Engl J Med* 1998; 339: 380–386.
6. European Medicines Agency. EPAR of Lartruvo (olaratumab), http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/004216/WC500216871.pdf (2017, accessed 1 July 2017).
7. GRADEpro GDT. GRADEpro Guideline Development Tool (developed by Evidence Prime, Inc.), McMaster University, http://www.gradepro.org (2015, accessed 1 July 2017).
8. Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336: 924–926.
9. GraphPad. Quickcalcs Section, http://graphpad.com/quickcalcs/ttest1/?Format=SEM (2016, accessed 1 July 2017).
10. Guyot P, Welton NJ, Ouwens JN, *et al.* Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. *Value Health* 2011; 14: 640–646.
11. Young R. Value-based cancer care. *N Engl J Med* 2015; 373: 2593–2595.