Journal of Hepatology 49 (2008) e672–e673

Journal of Hepatology

www.elsevier.com/locate/jhep

Journal Club

Special Section Editors: Peter R. Galle, Peter L.M. Jansen, Francesco Negro

Choosing the best endpoint ☆

Erik Christensen*

Department of Medicine I, Division of Gastroenterology, Bispebjerg University Hospital, Bispebjerg Bakke 23, DK-2400 Copenhagen NV, Denmark

Link to free abstract at: http://jnci.oxfordjournals.org/cgi/content/abstract/100/10/698

Like everything else, a disease develops and progresses in time. A valid recording and description of the course of disease is important both for evaluation of prognosis and effect of therapy. However, a detailed description of the course of disease can be very complex. A simpler but still valid description is obtained by recording the time from a defined starting point, e.g. the time of diagnosis or inclusion into a controlled clinical trial to the occurrence of an event or *endpoint* of interest, traditionally death as in *survival analysis*. Other endpoints like recurrence of neoplasm, occurrence of complications, alpha-feto-protein above a certain limit, etc. may be used, although the precise time of occurrence may be difficult to record. If in a patient the endpoint has not occurred before the end of the study, the observation is incomplete or "censored" for that patient.

The summarized result for a group of patients is the "survival" curve, which shows the cumulative "survival" probability (the cumulative probability of not having the endpoint) with time. Curves from various groups can be compared statistically using the logrank test and the Cox regression model [1] can be used to adjust for imbalance in prognostic variables between treatment groups and to identify patient characteristics associated with a beneficial or even harmful effect of therapy. This can serve to identify responders and non-responders to the therapy. The number of endpoints determines the power of statistical comparisons. Thus if many observations are incomplete or censored, the statistical power may be seriously reduced.

An ideal endpoint should have the following qualities: it should be *relevant* (central to the disease process), *accurate* (measure what it is purporting to measure), *precise* (measure the underlying quality with minimal error and minimal variability), *reliable* (a measurement made twice should produce the same results), and *assessable* within a relatively short time period. Although time to death or survival time may be free of bias, it may not be ideal because of the relatively long observation time needed if the number of censored observations should not be too large. In hepatocellular carcinoma (HCC) the severity of the underlying cirrhosis may influence the overall survival, and it may sometimes be difficult to ascertain whether a given patient died from HCC or cirrhosis [2]. Thus in HCC cancer-specific death may not be a very reliable endpoint. Therefore, so-called surrogate endpoints would be necessary to monitor the course of disease with relatively short time intervals. However, surrogate endpoints must be biologically plausible, reflect clinical severity, and correlate with the true outcome, i.e. survival time. They should also be responsive, such that the measurement changes in the anticipated direction either with disease progression or with the intervention.

Llovet et al. [2] list a total of nine usable endpoints in HCC, the most important being response rate, time to recurrence and time to progression (see Box 1).

The response rate should be based on the size of viable tumour tissue using contrast-enhanced radiological imaging in the arterial phase. Response may be complete (disappearance) or partial (30% linear decrease in tumour lesions or more). Non-response is either pro-

☆ The author declares that he does not have anything to disclose regarding funding from industries or conflict of interest with respect to this manuscript.

* Tel.: +45 3531 2854; fax: +45 3531 3556.
E-mail address: ec05@bbh.regionh.dk.

doi:10.1016/j.jhep.2008.07.012

gressive disease (20% linear increase in lesions or more) or stable disease (no increase or decrease in lesions).

Time to recurrence of neoplasm is defined as radiological demonstration of "new" (previously undetected) tumour tissue after resection or local ablation. A reliable recording of this endpoint is dependent on radiological imaging being performed with similar, reasonably short intervals in all the patients. Otherwise endpoint detection may be imprecise being delayed in some patients. This may weaken analysis and in extreme cases invalidate results.

Time to progression of neoplasm is also dependent on radiological imaging being performed with similar, reasonably short intervals in all the patients. This endpoint is also dependent on standardized criteria for what changes constitute a progression [2].

These three endpoints depend on defined criteria for how the radiological imaging results should be interpreted. The criteria serve to transform a quantitative variable (size of nodules) to a qualitative variable suitable for survival analysis methodology. However, the transformation from a quantitative to a qualitative variable will mean loss of information. Furthermore, by using survival analysis, the power may be decreased if there are many censored observations, and the delay between the radiological imaging investigations may imply imprecise endpoint detection as mentioned above.

It would be relevant to use the radiological imaging results directly without transformation. Estimation as precisely as possible of the *total volume* of active neoplastic tissue should be expected to provide the best information about the tumour burden of the patient. There has been a marked development in 3-dimensional (3D) imaging techniques in recent years [3–7]. An advantage of 3D volumetric measurement over 2D (or 1D) measurement is that when performing the subsequent scan, there is no need to select matching images on the follow-up study, since total volume is measured, which makes the volumetric measurements independent of slice registration. In lung cancer such techniques can be used to measure nodule volume accurately to within ±3% [5]. However, such precise results may not be extrapolated directly to HCC.

By measuring the total volume of active neoplastic tissue at certain defined time intervals in all the patients, the development of the disease may be described in more detail. Comparison of treatments in controlled clinical trials could then be performed using simple standard methods for quantitative variables. Using this approach the aforementioned problems of survival analysis could be avoided. Concentrating on one variable being central to the disease process could improve evaluation of the course and outcome of the disease, facilitate comparison of therapies and render many of the other, weaker surrogate endpoints less interesting. It would also be easier to compare results obtained in various centres.

References

[1] Christensen E. Multivariate survival analysis using Cox's regression model. Hepatology 1987;7:1346–1358.
[2] Llovet JM, Di Bisceglie AM, Bruix J, Kramer BS, Lencioni R, Zhu AX, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. J Natl Cancer Inst 2008;100:698–711.
[3] Badea R, Socaciu M, Lupşor M, Moşteanu O, Pop T. Evaluating the liver tumors using three-dimensional ultrasonography. A pictorial essay. J Gastrointestin Liver Dis 2007;16:85–92.
[4] Atasoy C, Akyar S. Multidetector CT: contributions in liver imaging. Eur J Radiol 2004;52:2–17.
[5] Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. Radiology 2000;217:251–256.
[6] Kostis WJ, Yankelevitz DF, Reeves AP, Fluture SC, Henschke CI. Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up CT. Radiology 2004;231:446–452.
[7] Enquobahrie AA, Reeves AP, Yankelevitz DF, Henschke CI. Automated detection of small pulmonary nodules in whole lung CT scans. Acad Radiol 2007;14:579–593.