

## Oncology endpoints: An unexpected journey

Waseem Jugon, Roche Products Ltd., Welwyn Garden City, England  
Mijanur Rahman, Roche Products Ltd., Welwyn Garden City, England

### ABSTRACT

What is an endpoint? Why is it important? In a clinical trial, endpoints are the key measure to assess the clinical benefit of the treatment to patients. For a new starter, oncology endpoints can often be complex and confusing to understand, but they are critical when evaluating cancer therapies in clinical trials. The role of a programmer should not only be able to analyse and interpret the data correctly as per protocol, but have some knowledge of the therapeutic area. This concept of programming whilst evolving into a data scientist can be somewhat unexpected, but understanding the scientific background enables a programmer to put an endpoint into context of a study. This paper describes the journey from data collection, to the types of statistical tests implemented and why they are important in the translation of assessing patient benefit in oncology trials.

### INTRODUCTION

The role of a programmer is constantly evolving and moving forward programmers will most likely not be producing tables, figures and listings (TFLs). The authors have worked within several oncology projects for a number of years and experienced many opportunities for new ways of thinking, novel ways of conducting analyses and further expanding the author's knowledge within the therapeutic area. As programmers we cannot stand still, we need to broaden our skill set, to show that we can add value to project work. Working within the oncology field has greatly enhanced the authors' knowledge, not just technically, but by investing time in understanding the analyses within the context of the disease area. The information in this paper describes what a programmer new to oncology may find unexpected, but will aid the development and understanding that can be applied to any other therapeutic area.

### ONCOLOGY BACKGROUND

The term oncology means a branch of science that deals with tumours and cancers. The term "Onco" refers to bulk, mass or tumour whilst the term "logy" means study.

Oncology is the most significant therapeutic area in terms of drug industry R&D and represents approximately 38% of the clinical trials market. (Median-Technologies, 2011-2013) However many of these drugs do not make it to market leading to huge financial losses amongst pharmaceutical companies and new drugs may require long term follow up before approval is gained.

### WHAT IS CANCER?

Cancer is the leading cause of disease worldwide with estimates of approximately 12.7 million new cases occurring worldwide and 7.6 million deaths in 2008. (WHO, 2013)

There are various different types of cancers; however, they are all usually formed when a swelling of part of the body occurs, usually without inflammation, which is caused by an abnormal growth of tissue. The body is made up of millions of living cells; these cells have a tightly regulated process that controls their growth, division and death. When cells of the body at a particular place start to grow out of control, they may become cancerous. The cancer occurs when the cells do not die, but carry on growing and forming new or abnormal cells. This is known as cell mutation and can form tumours in the body.

There are different types of tumours; solid and non-solid tumours. Solid tumours are excess tissues made up of cancer cells. Leukaemia is a non-solid tumour resulting from defective blood cells. Those that are not, are referred to as benign tumours; and those that are, are referred to as malignant tumours. Benign tumours can often be removed and in most cases do not return. Cells in these types of tumours do not invade other organs or spread to other parts of the body but they can grow very large and press on healthy organs and tissue. Malignant tumours are the cancerous cells; which can invade nearby tissues and spread to other parts of the body.

## DIAGNOSIS OF CANCER

Medical professionals who specialise in cancer are referred to as oncologists. They help in diagnosis of the cancer and grading the aggressive nature of the cancer using certain tools or methods. The most important diagnosis tool remains the medical history of the patient. This is to observe any common symptoms such as, fatigue, weight, unexplained anaemia, fever of unknown origin etc. There are times when a malignancy is located by performing a physical examination. Otherwise, diagnostic methods are used such as biopsies, x-rays, CT or MRI scanning, PET scans, ultrasounds, blood tests etc. After identifying the cancer, the oncologist plans the therapy that is suitable for the patient. This could be surgery, chemotherapy, radiotherapy and other modalities.

## EVALUATION OF CANCER THERAPIES

How are cancer therapies evaluated? The same methods as used in diagnosis are often used to monitor the patient's condition and response to treatment. For example, specialist blood tests and laboratory counts are recorded for non-solid tumours.

For solid tumours, the investigator will evaluate the size of the tumour based on a set of guidelines. The most common set of guidelines now used is The Response Evaluation Criteria In Solid Tumours (RECIST). These are a set of published rules first published in February 2000 but revised in 2008 which define when cancer patients improve ("respond"), stay the same ("stable") or worsen ("progression") during treatments or therapies. The RECIST criteria are a topic of immense interest and are important to understand. This paper only gives a high level overview of the criteria. The scope of RECIST with respect to this paper is to understand the terminology and what the response means in terms of patient benefit. Further in depth reading is available in the references section.

The evaluations of tumours are often collected as lesions for the purpose of analyses. These can be Target or Non-Target; however the presence of New Lesions may occur at subsequent tumour assessments.

The table below denotes the terms and definitions associated with attributing a response to a tumour assessment.

| Response                        | Target Lesions                                                                                                                                                                   | Non-Target Lesions                                                                                                                                   |
|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Complete Response (CR)</b>   | Disappearance of all target lesions. All pathological lymph nodes must have decreased to <10 mm in short axis                                                                    | Disappearance of all non-target lesions. All lymph nodes must be non-pathological in size (<10 mm short axis). Normalization of tumour marker level. |
| <b>Partial Response (PR)</b>    | At least a 30% decrease in the sum of longest diameters (SLD) of target lesions, taking the baseline SLD as reference.                                                           | N/A                                                                                                                                                  |
| <b>Non-CR / Non-PD</b>          | N/A                                                                                                                                                                              | Persistence of one or more non-target lesion(s) and/or maintenance of tumour marker level above the normal limits                                    |
| <b>Stable Disease (SD)</b>      | Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD                                                                                         | N/A                                                                                                                                                  |
| <b>Progressive Disease (PD)</b> | SLD increased by at least 20% from the smallest value on study (including baseline, if that is the smallest) The SLD must also demonstrate an absolute increase of at least 5mm. | Unequivocal progression of existing non-target lesions.                                                                                              |

**Table 1 – Definitions are as per the RECIST 1.1 guidelines (Eisenhauer, et al., 2008)**

The lesions are assessed at regular intervals as per protocol and an overall response is derived at each tumour assessment by the investigator or independent reviewer or adjudicator. These are based on the readings and measurements of the tumour by applying a set of defined algorithms (usually RECIST guidelines).

It is important a programmer understands the algorithm used and what these test results mean to a patient. This can empower the programmer to question data cleanliness and challenge certain responses that may not look correct by looking further into the data. Response of tumour assessments and interpreting this information correctly is key in any oncology study in determining treatment effect and benefit. For example, Objective response rate and Progression free survival endpoints utilize this information to make decisions on the efficacy of the treatment.

## PhUSE 2013

### ONCOLOGY ENDPOINTS

An endpoint can be defined as a measure of evaluating cancer therapies. When performing clinical trials, the primary endpoint of a study must be able to provide a valid and reproducible measure of clinical and statistical benefit to the patient population being treated. Clinical endpoints are defined in the protocol of the clinical trial and the methods of interpreting the data of how to calculate these are contained within a statistical analyses plan.

*“As payers seek to keep pace with groundbreaking changes in the oncology arena, it is critical that they have a solid understanding of how oncology clinical trial endpoints can or should be used to guide decisions about the care that patients receive.” (Kogan & Haren, 2008)*

Just as payers need to have a solid understanding of clinical trial endpoints, as programmers, we also need to have a solid understanding and knowledge of the data and the scientific evidence of clinical benefit to the patient.

Some endpoints may be referred to as surrogate endpoint. A surrogate endpoint can be used as an indication to predict clinical benefit. These are becoming more advantageous to use in order to help get accelerated approval of the drug as events may occur earlier, it is less expensive and more importantly, can help bring beneficial novel treatments to patients quicker.

### COMMON TYPES OF ENDPOINTS USED IN ONCOLOGY STUDIES

| Statistical Endpoint                   | Definition                                                                                                                                                                                                                                                                                             | Limitation                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Overall survival (OS)</b>           | Time from randomization until death from any cause; most commonly used endpoint in phase 3 trials and trials for regulatory approval and is considered the gold standard used to determine patient benefit. The reason this endpoint is preferred is that it is not subject to any investigator bias.  | <ul style="list-style-type: none"> <li>Requires randomized trial with lengthy follow-up.</li> <li>Can be affected by subsequent therapies</li> </ul>                                                                                                                                                                                                                                              |
| <b>Progression-free survival (PFS)</b> | A surrogate endpoint for OS. Time from randomization to objective tumour progression or death. Unlike time to progression (TTP), PFS includes death from any cause as well as progression. Like TTP, it is unaffected by subsequent therapies. FDA prefers PFS rather than TTP as regulatory endpoint. | <ul style="list-style-type: none"> <li>Not statistically validated as surrogate for survival in all treatment settings</li> <li>Not precisely measured subject to assessment bias particularly in open-label studies</li> <li>Definitions vary among studies</li> <li>Frequent radiological or other assessments</li> <li>Involves balanced timing of assessments among treatment arms</li> </ul> |
| <b>Time to progression (TTP)</b>       | A surrogate endpoint for OS. Defined as time from randomization until objective tumour progression. Unlike PFS, it does not include deaths, but if most deaths are not cancer-related TTP can be acceptable endpoint. Like PFS, it is unaffected by subsequent therapies.                              |                                                                                                                                                                                                                                                                                                                                                                                                   |
| <b>Objective response rate (ORR)</b>   | A surrogate endpoint for OS. A proportion of patients with reduction in tumour size by a predefined amount (using standardized criteria, such as RECIST). Directly attributable to drug effect.                                                                                                        | <ul style="list-style-type: none"> <li>Not a direct measure of benefit</li> <li>Not a comprehensive measure of drug activity</li> <li>Only a subset of patients who benefit</li> </ul>                                                                                                                                                                                                            |
| <b>Patient Reported Outcomes (PRO)</b> | Patient-reported outcomes, such as quality of life (QOL), complement information from traditional endpoints, generating the patient's global assessment of the direct clinical benefit of a drug                                                                                                       | <ul style="list-style-type: none"> <li>Blinding often is difficult</li> <li>Data frequently are missing or incomplete</li> <li>Clinical significance of small changes is unknown</li> <li>Multiple analyses</li> <li>Lack of validated instruments</li> </ul>                                                                                                                                     |

**Table 2 Data and definitions of selected endpoints: (Genentech USA, The Ongoing Evolution Of Endpoints in Oncology, 2011) and (FDA, 2007)**

### NOVEL ENDPOINTS AND THE FUTURE

The introduction of surrogate endpoints has enabled accelerated approval for treatments to bring novel treatments to patients faster. There are many advantages to this; however, the acceptance of using an endpoint as a surrogate requires validation from many trials and usually meta-data analyses is performed to check if there is any correlation between a surrogate and the primary endpoint. In most cases in oncology, the primary endpoint is overall survival. Overall survival is still considered the gold standard in terms of determining treatment benefit, but if a surrogate endpoint can be used to gain accelerated regulatory approval, more novel therapies can reach patients faster. PFS has been an example of a surrogate that has been used as the primary endpoint for a novel targeted therapy treatment in 1<sup>st</sup> line metastatic breast cancer, Perjeta® (pertuzumab). By using PFS as opposed to OS as an endpoint, this has enabled the treatment to be made available to patients quicker.

*FDA 'Breakthrough Therapy Designation' is designed to expedite the development and review of medicines intended to treat serious and life threatening diseases and to help ensure people have access to them through FDA approval with a shorter review time. A priority review designation is granted to medicines that the FDA believes have the potential to provide significant improvements in the treatment, prevention or diagnosis of a disease.*

The FDA recently published a draft guidance document for the use of Pathological Complete Response (pCR) in Neoadjuvant Treatment of High Risk Early-Stage Breast Cancer (FDA., 2012). Roche has recently filed Perjeta® for the use of the treatment in patients with early breast cancer in a high risk population, using pCR as the endpoint of interest. This followed the proposed new FDA pathway designed to more quickly bring promising medicines to people with earlier stages of breast cancer, where treatment may have a greater impact.

**Pathological complete response (pCR)** — There is no definitive definition of what can be considered pCR, however the FDA defines pCR as 'the absence of any residual invasive cancer on haematoxylin and eosin evaluation of the resected breast specimen and all sampled ipsilateral lymph nodes following completion of neoadjuvant systemic therapy'.

In simple terms, following neoadjuvant therapy, the patient is a responder if there is no evidence of disease in the breast or lymph nodes as examined by a pathologist.

Following the recent filing an advisory board was consulted with regards to whether or not the drug should be considered for regulatory approval. A recent press release from Roche stated:

*"Food and Drug Administration's (FDA) Oncologic Drugs Advisory Committee (ODAC) voted 13 to 0, with one abstention, in favour of recommending accelerated approval of a Perjeta® (pertuzumab) regimen for neoadjuvant treatment (use before surgery) in people with high-risk HER2-positive early stage breast cancer. The FDA will make a decision on whether or not to approve Perjeta for this use by 31 October, 2013. If approved, the Perjeta regimen will be the first neoadjuvant breast cancer treatment approved in the U.S. and the first treatment approved based on pathological complete response (pCR) data, meaning there is no tumour tissue detectable at the time of surgery."* (Roche., 2013)

Should approval be granted, this would be a major breakthrough for treatment in breast cancer and achieve a common goal of bringing treatment to patients quicker and safer. As programmers, understanding new data and the impact it has on patients' lives is critical. If we understand the data, we can question the data, provide exploratory analyses, spot trends and focus on subgroups. The Perjeta® filing for use in neoadjuvant treatment is an example of having to learn new science, understand new data points, and put this into the context of how much a patient can benefit from receiving treatment in the neoadjuvant setting and having confidence in the endpoint derived.

### DATA SCIENCE

The term data science has been around for decades and its definition varies depending on the job at hand. Data science is the integration of methods from statistics, computer science, and other fields for gaining insights from data. In practice, data science encompasses an iterative process of data harvesting, cleaning, analysis and visualization, and implementation. (Reveltyx, 2012) In the case of a statistical programmer, keeping up to date with technology, techniques, tools and even the types of data is a discipline that is required. The industry is changing and programmers have to adapt with the changes moving forwards. The biggest part of a programmers role will not be creating TLG's, but will be more non clinical trial reporting such as granular checking of the data, understanding of the data with scientific context, learning how to mine large amounts of data, perform meta-analyses and visualize data in a way that tells a story.

Understanding and grasping knowledge of clinical endpoints requires the skills mentioned. Not all the skills may fit within a programmers comfort zone, for example a programmer with only a computing background may find the

concept of some statistics or understanding the therapeutic area more difficult than others. Up-skilling to add value to the job we already do can make a significant difference to the contribution we make in decisions for streamlined and targeted analyses. A programmer can feel empowered to make decisions and have the confidence to sit down with a data manager, statistician or scientist to ensure the quality of work delivered and the meaning of the work delivered makes sense. Tools such as SAS®, R and Spotfire are amongst a selection of packages that can aid a programmer in exploring data not only to help answer key questions and also put forward new ideas.

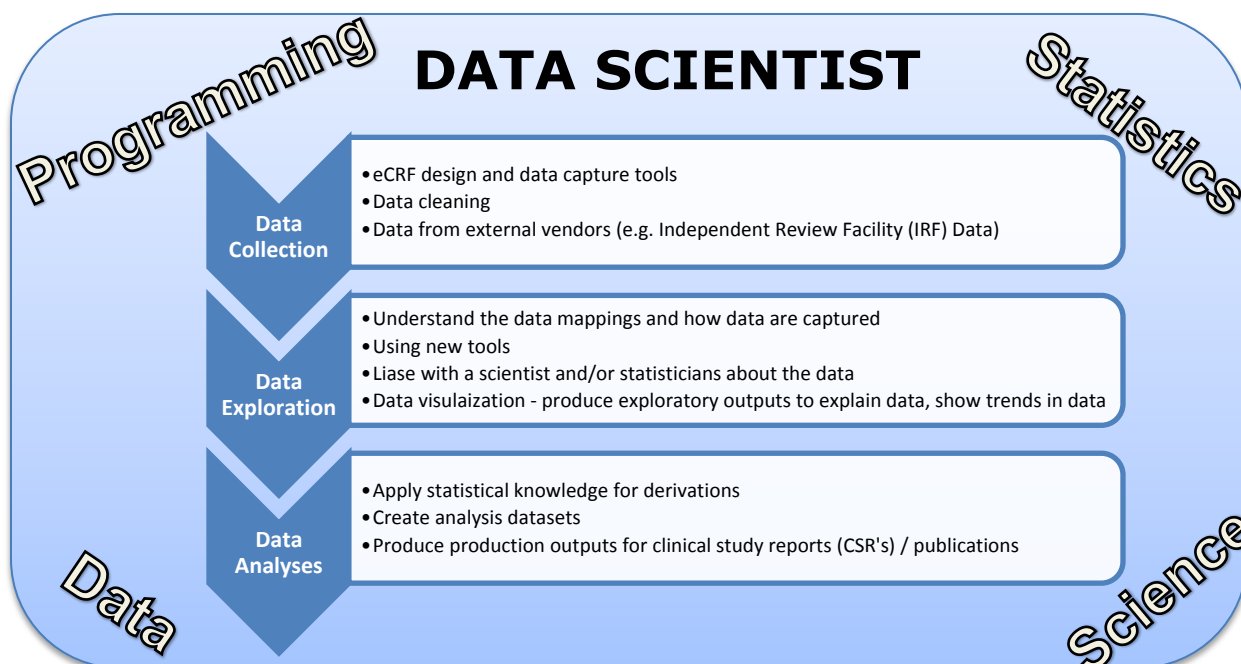


Figure 1 – A view of the different tasks a programmer can undertake and the overlap of expertise that may be required

A simple example of looking into the data to understand data used for an endpoint could be checking the correctness of an investigator assessed PD. For example, if we receive the investigator response directly from the CRF, how do we know this is correct? How do we know this is not a data entry error? Could we be over reporting or under reporting events? A programmer should ask these questions and can look into the data if they know the science, have knowledge of the RECIST criteria and the effect it has on the overall endpoint.

| Group ID | Visit Name               | Character Result/Finding in Std Format | Standard |
|----------|--------------------------|----------------------------------------|----------|
| TARGET   | SCREENING                | 51                                     | mm       |
| TARGET   | TUMOR ASSESSMENT WEEK 6  | 52                                     | mm       |
| TARGET   | TUMOR ASSESSMENT WEEK 12 | 56                                     | mm       |
| TARGET   | TUMOR ASSESSMENT WEEK 18 | 56                                     | mm       |

Figure 2 – Data for a patients Target lesion at screening and subsequent visits.

| Response Assessment Name | Visit Name               | Response Assessment Result in Std Format |
|--------------------------|--------------------------|------------------------------------------|
| Overall Response         | TUMOR ASSESSMENT WEEK 6  | SD                                       |
| Overall Response         | TUMOR ASSESSMENT WEEK 12 | SD                                       |
| Overall Response         | TUMOR ASSESSMENT WEEK 18 | PD                                       |

Figure 3 – Data for investigator assessed overall response for post baseline tumour assessments.

The patient had a target lesion which is measured up to week 18 (Figure 2). The target lesion has **increased in size by 5mm** from screening to 56mm. The investigator has marked this patient **at week 18 as having a PD** (Figure 3). Is this correct? On face value, the tumour has increased but as per RECIST v1.1 definition, **not by 20%**. 5mm is **only a 9.8% change**. More investigation is required. The only way that this patient can now be marked as a PD is if the patient had any **new lesions observed at the week 18 assessment**.

## PhUSE 2013

The data shows **no evidence of new lesions**; this can then be deemed an issue to query with science and data managers to ensure this information is correct. By understanding the data and the criteria used to determine information for a **critical endpoint (PFS)**, if this scenario had not been checked, it could potentially lead to **over reporting of events** making the treatment look less efficacious than it actually could be.

### STATISTICAL ANALYSES

There are a number of different statistical tests presented in outputs when trying to interpret patient benefit. Here are a few key common statistical terms and definitions that are useful to aid a programmer to interpret the meaning of a clinical endpoint:

| Statistic                       | Definition                                                                                                                                                                                                                                                                                                    |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>N</b>                        | Number of subjects in the study population                                                                                                                                                                                                                                                                    |
| <b>n</b>                        | Number of subjects in a sample from the study population.                                                                                                                                                                                                                                                     |
| <b>Mean</b>                     | Sum of the observed values divided by the number of observations. The most common value found in the distribution or sample.                                                                                                                                                                                  |
| <b>Median</b>                   | The middle observation when the observed values are ranked from smallest to largest. This is the numeric value separating the higher half of a sample.                                                                                                                                                        |
| <b>Range</b>                    | The difference between the max and min results in a set of figures. Showing the spread of the data.                                                                                                                                                                                                           |
| <b>Standard Deviation</b>       | A measure of the spread of scores away from the mean.                                                                                                                                                                                                                                                         |
| <b>Confidence Interval (CI)</b> | Range of values within which we are fairly confident the true population value lies. For example, a 95% CI means that we can be 95% confident that the population value lies within those limits.                                                                                                             |
| <b>P-Value</b>                  | The probability of observing the estimated treatment difference, or one more extreme, if the treatments are equally effective.                                                                                                                                                                                |
| <b>Response Rate</b>            | The number of patients who responded to treatment expressed as a proportion of the total number of patients in the treatment arm                                                                                                                                                                              |
| <b>Censoring</b>                | Censoring occurs when patients included in the study have not experienced an event of interest in survival analysis. This also includes patients who have been lost to follow up and their exact survival time is unknown.                                                                                    |
| <b>Hazard Ratio (HR)</b>        | The relevant risk of experiencing an event being measured (e.g. death) between 2 groups. A HR of 1 indicates no difference. A HR of < 1 typically means there was reduced risk in one of the treatment arms (usually experimental arm) and a HR > 1 indicates an increased risk in one of the treatment arms. |
| <b>Odds Ratio (OR)</b>          | This is the ratio of an event happening compared to an event not happening in the sampled population.                                                                                                                                                                                                         |

Table 3 – Table of key statistical definitions (Machin, Campbell, & Walters, 2007) and (Harris & Taylor, 2008)

### SURVIVAL ANALYSIS

In clinical trials, the effect of an intervention is evaluated by measuring **the number of patients survived after that intervention over a period of time**. The time starting from a defined point (e.g randomization) to the occurrence of a given event (e.g. progression of disease or death), is referred to as survival time and the analyses group data as survival analyses. Certain situations can occur in a study such as patients refuse to remain in the study, or loss of contact with the patient, or when some patients may not experience death before the end of the study; even though they would have experienced death if the patient continued. For these patients we may only have partial information

but they are not ignored as they provide some information about survival. All these situations are labelled as censored observations. The Kaplan-Meier estimate is one of the simplest ways of computing the survival over time in spite of these difficulties.

## KAPLAN-MEIER CURVE

The Kaplan- Meier (KM) curve is a graphical representation of a **Time to Event** analysis showing when a patient reaches a trials survival endpoint. It can be summarised by observing the median survival. Median survival is the measure of how long patients will live on average with a certain disease or treatment, and corresponds to the point on the KM curve where the survival probability equals 0.5. The KM estimate describes the probability of surviving in a given length of time while considering time in many small intervals depicted as steps, occurring at the time of each new event. It involves calculating probabilities of occurrence of event at a certain time point, based on the number of patients at risk of the event at the start of the time interval and multiplying these successive probabilities by any earlier computed probabilities to get the final estimate. It also gives an approximate indication of the survival as well as the prognosis of a group of patients with cancer.

## LOG RANK TEST

The Log rank test is solely a hypothesis test and provides no direct information on how different the treatment groups are but can be to compare the KM survival distributions between 2 groups. The test assumes that there is no difference between the survival curves. This test is acceptable to use with KM distributions and accounts for censoring. This test is commonly used in clinical trials to give an indication of the efficacy of a new treatment to standard of care. The P-value can tell us if the difference between the survival distributions is statistically significant. This test may be seen as the Mantel-Cox test or as a time stratified Cochran-Mantel-Haenszel test. The size of the treatment difference can be estimated by calculation of a HR and associated confidence interval, which is used as a measure of relative survival between each group.

## COX REGRESSION

The Cox regression model provides us with estimates of the effect that different factors have on the time until the end event. For example, the effect that age, weight, sex etc... might have on the time to event analysis. The main aim in this model is to understand the HR. The HR takes into account of the number and timing events, and the time until last follow-up for each patient who has not experienced an event i.e. has been censored..

There are further in-depth readings available on survival analysis in the references; this paper purely provides a high level overview of these widely used statistical methods used in oncology to assess efficacy.

## INTERPRETING RESULTS TO ASSESS PATIENT BENEFIT

Once summary tables have been produced, it is important to understand the output that has been created. Ask yourself the question, "Does this output make sense?" Understanding the science and statistics can allow the programmer to evaluate and perhaps even question an output before formally releasing the output for QC and functional team review. This section gives a few examples of typical oncology outputs containing key endpoints and the translation into assessing meaningful patient benefit.

## TIME TO PROGRESSION

|                               | Control Arm<br>(N=406) | Experimental Arm<br>(N=402) |   |
|-------------------------------|------------------------|-----------------------------|---|
| Patients included in analysis | 406 (100.0 %)          | 402 (100.0 %)               | 1 |
| Patients with event           | 242 ( 59.6 %)          | 191 ( 47.5 %)               |   |
| Patients without event        | 164 ( 40.4 %)          | 211 ( 52.5 %)               |   |
| Time to event(Months)         |                        |                             | 2 |
| Median                        | 12.4                   | 18.5                        |   |
| 95% CI for Median             | [10;13]                | [15;23]                     |   |
| 25% and 75%-ile               | 6;25                   | 9; .                        |   |
| Range                         | 0 to 33                | 0 to 34                     |   |
| p-value (Log-Rank test)       | <.0001                 |                             | 3 |
| Hazard Ratio                  | 0.63                   |                             |   |
| 95% CI                        | [0.52;0.76]            |                             |   |
| 1 year duration               |                        |                             | 4 |
| Patients remaining at risk    | 161                    | 211                         |   |
| Event Free Rate               | 0.51                   | 0.65                        |   |
| 95% CI for Rate               | [0.46;0.56]            | [0.60;0.70]                 |   |

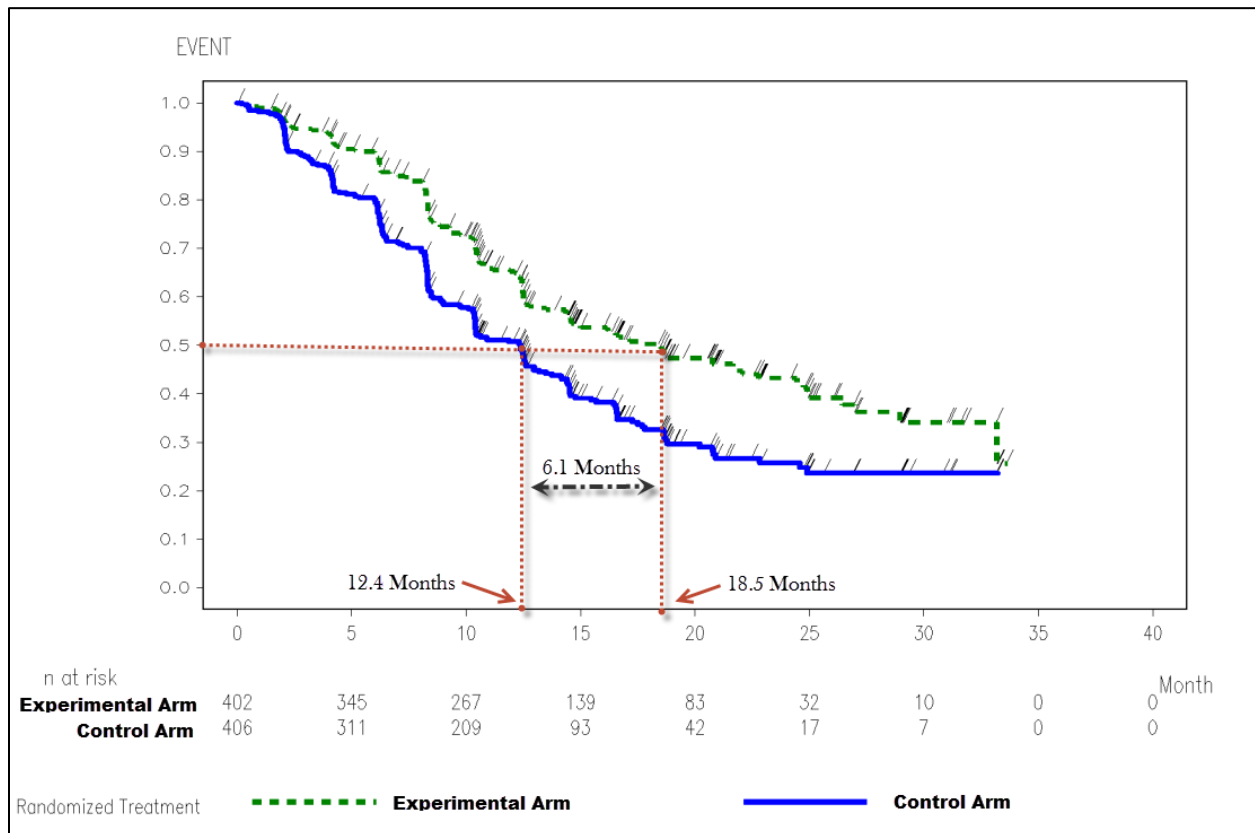
Figure 4

## PhUSE 2013

**Figure 4** shows summary statistics for a PFS endpoint. The output is interpreted as follows in order to assess the benefit of the treatment to the patient:

|                              |                                                                                                                                                                                                                                                                                                         |
|------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1. Responders</b>         | Nearly <b>60% of patients in the control arm</b> were deemed to have had an event (In this example progressed or died as per PFS definition) compared to approximately <b>48% in the experimental arm</b> . Patients with no PFS event also include those patients who have been censored.              |
| <b>2. Time to Event</b>      | The median time of PFS in the <b>control arm was 12.4 months</b> . The experimental arm shows an <b>improvement of PFS by approximately 6.1 months</b> . A P-Value calculated using a log rank test of <b>less than 0.0001 indicates that this is statistically significant</b> .                       |
| <b>3. Hazard Ratio</b>       | A Hazard ratio of <b>0.63</b> indicates that there is <b>37% lower risk of a PFS event</b> occurring in the experimental arm compared to the control arm.                                                                                                                                               |
| <b>4. Truncated Analyses</b> | At 1 year duration into the study, <b>211 of the 402 patients have not progressed or died</b> in the experimental arm. This is significantly higher than <b>only 161 patients remaining at risk in the control arm</b> . Therefore more patients have not experienced an event in the experimental arm. |

### KAPLAN MEIER SURVIVAL CURVE FOR TIME TO PROGRESSION ANALYSES



**Figure 5**

**Figure 5** shows a survival curve for the same PFS endpoint summarised in **Figure 4**. The benefit of survival curves is to graphically show the trend in data and immediately see the impact of the trial treatment between the 2 treatment groups. It clearly tells a story that at the point where 50% of events occurred in each arm, the median survival time is greater in the experimental arm. The values are annotated on the figure for clarity. The "tick" marks on each line represent any censored observations.

## OBJECTIVE RESPONSE RATES ANALYSES

|                                         | Control Arm<br>(N=371) | Experimental Arm<br>(N=367) |   |
|-----------------------------------------|------------------------|-----------------------------|---|
| Responders                              | 253 ( 68.2 %)          | 284 ( 77.4 %)               | 1 |
| Non-Responders                          | 118 ( 31.8 %)          | 83 ( 22.6 %)                |   |
| 95% CI for Objective Response Rate      | [ 63.2; 72.9]          | [ 72.8; 81.6]               |   |
| Difference in Response Rates            | 9.19                   |                             | 2 |
| 95% CI for Difference in Response Rates | [ 2.7; 15.7]           |                             |   |
| p-Value                                 | 0.0049                 |                             |   |
| Odds Ratio                              | 1.60                   |                             | 3 |
| 95% CI for Odds Ratio                   | [1.15; 2.22]           |                             |   |
| Complete response (CR)                  | 22 ( 5.9 %)            | 25 ( 6.8 %)                 |   |
| 95% CI for complete response rate       | [ 3.8; 8.8]            | [ 4.5; 9.9]                 | 4 |
| Partial Response (PR)                   | 231 ( 62.3 %)          | 259 ( 70.6 %)               |   |
| 95% CI for partial response rate        | [ 57.1; 67.2]          | [ 65.6; 75.2]               |   |
| Stable disease (SD)                     | 78 ( 21.0 %)           | 52 ( 14.2 %)                |   |
| 95% CI for stable disease rate          | [ 17.0; 25.5]          | [ 10.8; 18.2]               |   |
| Progressive disease (PD)                | 23 ( 6.2 %)            | 14 ( 3.8 %)                 |   |
| 95% CI for progressive disease rate     | [ 4.0; 9.2]            | [ 2.1; 6.3]                 |   |

Figure 6

The summary statistics and response rates for Objective Response are shown in **Figure 6**. The output is interpreted as follows in order to assess the benefit of the treatment to the patient:

|                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1. Responders</b>                 | According to RECIST, a response can be defined as a patient's lesion being <b>reduced in size by at least 30%</b> . This is identified as either having a <b>CR or PR</b> . Approximately <b>77% of the patients respond</b> to treatment in the <b>experimental arm</b> in comparison to approximately <b>68% in the control arm</b> .                                                                                                                                                                  |
| <b>2. Treatment Group comparison</b> | The difference in the response rates between the 2 treatment arms is approximately <b>9.2 %</b> . The <b>P-value of about 0.005</b> implies that there is a <b>statistically significant difference</b> between the 2 treatment groups.                                                                                                                                                                                                                                                                  |
| <b>3. Odds Ratio</b>                 | An Odds ratio of <b>1.60</b> indicates that the likelihood of achieving a response is 1.6 times higher in the experimental arm compared to the control arm. This is derived by looking at the proportion of patients who responded divided by the number of non-responders for each treatment arm i.e. When we divide the result of the <b>proportion in the experimental arm by the proportion in the control arm</b> , that is <b>(284/83) / (253/118)</b> , the result is approximately <b>1.60</b> . |
| <b>4. Response Rate</b>              | The number of responses for each observed tumour response is calculated as a percentage out of the population N. The associated 95% confidence intervals are also calculated. The combined totals for CR and PR will equal the number of responders in the first block.                                                                                                                                                                                                                                  |

## FOREST PLOT ANALYSES

Forest plots are extremely useful when it comes to displaying or trying to identify the treatment effect in different subgroups of patients. This can be quite useful when exploring data and help identify any key subgroup populations that may require further analyses and also give an indication when designing new trials and new analyses, if there is a treatment effect in a particular subgroup of patients.

Forest plots are usually created for exploratory analyses as many clinical trials are not always designed to show treatment benefits in many different subgroups. The plot shows the corresponding magnitude of benefit and confidence limits of each subgroup analysed.

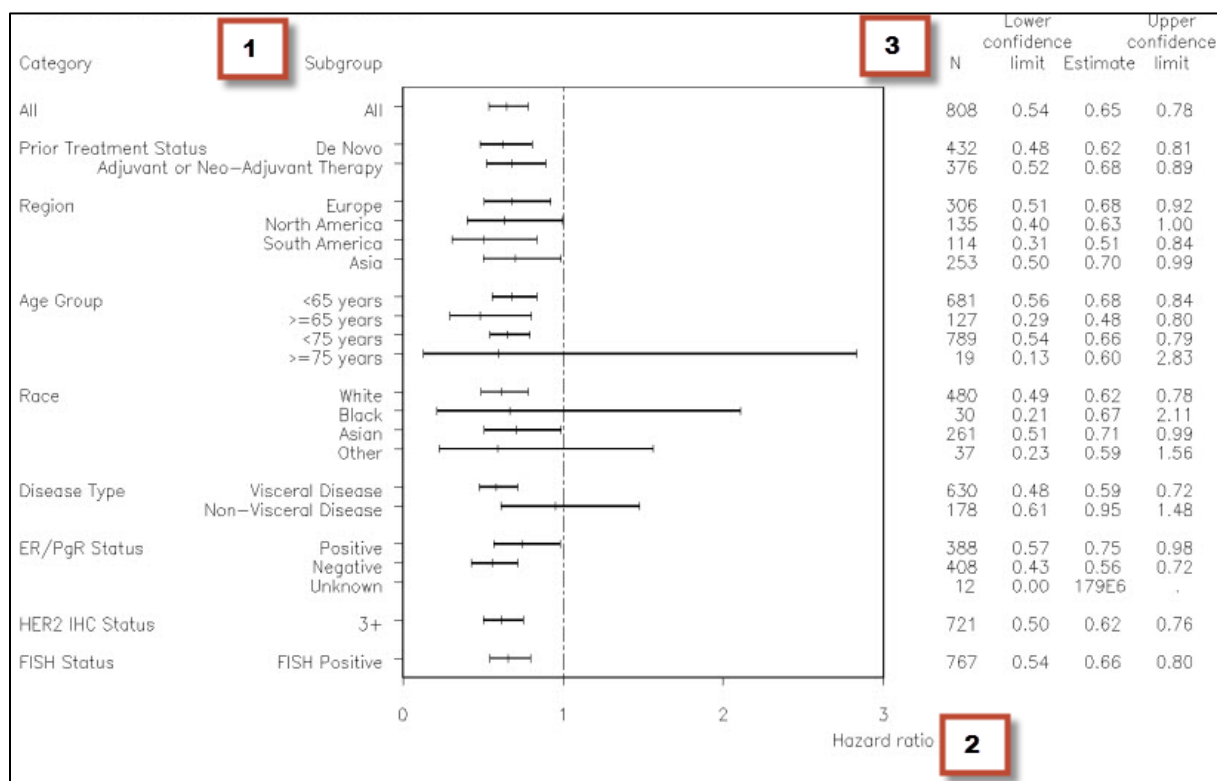


Figure 7

Figure 7 shows a forest plot for the overall population of the trial and various different subgroups. How is this interpreted?

|                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1. Category and Subgroups</b>                    | To the left of the graph are the desired categories of interest with the values associated for each subgroup.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <b>2. Hazard Ratio</b>                              | The horizontal axis is the HR. This is the magnitude of treatment benefit to the patient. The dashed line at a HR of 1, is the point where there is no difference in effect between treatments. <b>Any HR that is less than 1, favours the experimental treatment arm and any HR &gt; 1 favours the control arm.</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>3. Magnitude of Benefit and Confidence Level</b> | <p>To the right of the graph, the population sizes of each subgroup are denoted by N. This can show the relative size of the subgroup. The estimate (HR) is also displayed alongside its confidence intervals. The estimate is plotted on the graph denoted by the midpoint tick across the line plotted and the lower limit as the very left hand side tick and upper limit as the very right hand side tick.</p> <p>From this plot we can see that the <b>magnitude of benefit quite distinctively favours the experimental arm in all subgroups.</b></p> <p>However, there <b>are 4 results that seem to favour the experimental arm but the upper CI limit falls into favouring the control arm.</b> Looking at N, the sample size for the subgroup of interest is fairly small and <b>the confidence of these results is relatively low.</b></p> |

### CONCLUSION

For a programmer who is new to oncology and to the pharmaceutical industry, it is important to understand and interpret the analyses of the work done. A programmer's responsibility should not only be to produce TLG's from a set of specifications, but understand the therapeutic area, have a clear understanding of the key data points collected and the assumptions behind the statistical analyses done.

This empowers the programmer to challenge any analyses done and understand the key message that is being delivered to stakeholders. This may be out of the programmers comfort zone, however the benefits in the long term will mean that a programmer understands the journey from data collection, through to reporting.

A programmer may have to up skill to use a variety of new tools available in order to perform additional non-clinical reporting tasks quickly and more efficiently. Programmers are at the forefront of producing the analyses and exploring the data, so it is essential that the end product of all the work done can be accurately and meaningfully translated into describing patient benefit.

### REFERENCES & RECOMMENDED READING

- Eisenhauer, E., Therasse, P., Bogaerts, J., Schwartz, L., Sargent, D., Ford, R., et al. (2008, October 28). *New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)*. Retrieved 2013, from EUROPEAN JOURNAL OF CANCER: <http://www.eortc.be/Recist/documents/RECISTGuidelines.pdf>
- ELLENBERG, S. S., & HAMILTON, J. M. (1989). SURROGATE ENDPOINTS IN CLINICAL TRIALS: *CANCER. STATISTICS IN MEDICINE* (Vol. 8), 9.
- FDA. (2007). *Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics*. Retrieved 2013, from FDA: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf>
- FDA. (2012, May). *Guidance for Industry Pathologic Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval*. Retrieved 2013, from FDA: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM305501.pdf>
- Genentech USA, I. (2011). *The Ongoing Evolution Of Endpoints in Oncology*. Retrieved from NAMCP Medical Directors: <http://www.namcp.org/institutes/cancer/Oncology%20Endpoints.pdf>
- Genentech USA, I. (2013). *Analyzing the results of clinical trials: Biostatistics*. Retrieved from Bioncology: <http://www.bioncology.com/clinical-trials/clinical-endpoints/biostatistics>
- Harris, M., & Taylor, G. (2008). *Medical Statistics Made Easy*. Scion Publishing Limited.
- Kogan, A. J., & Haren, M. (2008, May). *Translating Cancer Trial Endpoints Into the Language of Managed Care*. Retrieved 2013, from Biotechnology Healthcare: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2651701/>
- Machin, D., Campbell, M. J., & Walters, S. J. (2007). *Medical Statistics: A Textbook for The Health Sciences*. John Wiley & Sons Ltd.
- Median-Technologies. (2011-2013). *Median Images are key in oncology*. Retrieved 2013, from Median Technologies: <http://www.mediantechologies.com/environment>
- Oncology, A. (2009). *CLINICAL DRUG TRIALS IN CANCER: STUDY ENDPOINTS – WHAT DO THEY REALLY MEAN?* Retrieved 2013, from PR Newswire: <http://multivu.prnewswire.com/mnr/astrazeneca/38570/docs/38570-ClinicalTrialsEndpointFactsheet.pdf>
- Pilz, L., Manegold, C., & Schmid-Bindert, G. (2012, March). *Statistical considerations and endpoints for clinical lung cancer studies: can progression free survival (PFS) substitute overall survival (OS) as a valid endpoint in clinical trials for advanced non-small-cell lung cancer?* Retrieved 2013, from Translational Lung Cancer Research: <http://www.tlcr.org/article/view/287/599>
- Revelytix. (2012). *What is Data Science?* Retrieved 2013, from Revelytix: <http://www.revelytix.com/?q=content/what-data-science-0>
- Roche. (2013, September 13). *FDA Advisory Committee recommends accelerated approval of Roche's Perjeta for neoadjuvant use in HER2-positive early stage breast cancer*. Retrieved from Roche: [http://www.roche.com/investors/ir\\_update/inv-update-2013-09-13.htm](http://www.roche.com/investors/ir_update/inv-update-2013-09-13.htm)
- WHO. (2013). *World Cancer Day 2013*. Retrieved 2013, from World Health Organisation: <http://www.who.int/cancer/en/>

### ACKNOWLEDGMENTS

The views expressed in this paper are those of the authors and do not necessarily reflect those of Roche or any other organisation that may be cited for reference in this paper.

The authors would like to thank Karen Rowe, Vijay Reddi, Chris Leigh, Hinal Patel, Hiren Naygandhi, Emma Clark, Richard Houghton, Paul Grimsey, Connor Steer, Dr Graham Ross, Riaz Jugon and Imran Jugon for review, feedback and support in writing this paper.

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

**Waseem Jugon**

Roche Products Ltd.  
Shire Park, Falcon Way  
Welwyn Garden City AL7 1TW  
Work Phone: 01707 366150  
Email: [waseem.jugon@roche.com](mailto:waseem.jugon@roche.com)

**Mijanur Rahman**

Roche Products Ltd.  
Shire Park, Falcon Way  
Welwyn Garden City AL7 1TW  
Work Phone: 01707 366531  
Email: [mijanur.rahman@roche.com](mailto:mijanur.rahman@roche.com)

Brand and product names are trademarks of their respective companies.