In the header we have the title “

|  |
| --- |
| **Figure 14.2.12.2b. Kaplan-Meier Plot of Overall Survival (Cohort 2: KTM-Z) (mITT Analysis Set, N = 14)”** |

Title always starts with words Table, Figure or Listing, followed by identifier numbers.

Step one is to capture the whole title in the document.

From the title we extract few predefined variables if available.

4 main categories of data we want to extract are in **Bold**:

**Cohort** (In this case Cohort 2, but can be)

* Cohort 1
* Cohort 2
* Cohort 3

**Analysis set** (we can see from this title the analysis set is mITT but it can be any of the following)

* mITT
* IAS or Inferential Analysis Set
* SAS or Safety Analysis Set
* FAS or Full Analysis Set

**Data cutoff date** (this case 31 DEC 2019)

This can be any date but always in the same format after “Data cutoff date = “

**Endpoint**

* Best overall response
* duration of response
* objective response rate
* Overall Survival
* Progression-Free Survival
* Time
* Safety