Appendix C. Statistical terms

Statistical terms	Definition
Age- specific incidence rate (AR)	The incidence rate of colorectal cancer was calculated for specified five-year age groups, by dividing the number of colorectal cancer cases in the five-year age group by the Sabah population in that particular age group and multiplying by 100 000. **AR = N / P x 100,000** **N – Number of new cancers occurring in the I age group** **P – Person-years at risk in the I age group**
	This was defined as the sum of age-specific rates of colorectal cancer in Sabah weighting to the world standard population ex pressed per 100 000. **ASR = Σ (Ari x Pi.std) / total world standard population** **Ari – Age specific rate in the ith age class** **Pi.std – The number in the ith age class of world standard population**
Age specific mortality rate	This was the mortality rate of colorectal cancer in Sabah in a specified five- year age group, and was calculated by dividing the number of colorectal cancer deaths in the five-year age group by the Sabah

	population in the particular age group and multiply by 100 000.
Age standardized mortality rate	This was defined as the sum of age-specific mortality rates of colorectal cancer in Sabah weighted to the world standard population expressed per 100 000.
Chi Square test	This is a non-parametric test used to compare more than two variables (age category below and above 50 years old with ethnic group indigenous and Chinese) from a randomly selected data.
Confidence Interval of the ASR (95% CI)	The confidence interval of the ASR (95% CI) is the range in which the true age-standardised rate is expected to fall, subject to some random variation. This is conventionally stated as age standardized rate plus or minus 1.96 multiplied by standard error.
Frequency	The frequency is the number of colorectal cancer cases in this study from January 2012 until December 2016. It is also used to depict the sum of each variable in this study pertaining to demographics, clinical and treatment characteristics.
Interquartile range	The interquartile range is the measure of variability of the median age of colorectal cancer cases based on diving the data into quartiles. The data for age of the colorectal cancer cases were divided into four equal parts and expressed in ranges.

Relative risk	This is the measure of colorectal cancer risk in one group (indigenous) compared another group (Chinese). It was calculated by diving the percentage of colorectal cancer cases in the indigenous group below the age of 50 years with the percentage of colorectal cancer cases in the Chinese group below the age of 50 years. **Relative risk 1: No difference** **Relative risk >1: Increased risk** **Relative risk < 1: Reduced risk**
Mean	The mean is the measure of central tendency of the age of the colorectal cancer cases in which is calculated by diving the sum the age over the frequency of the cases. The dataset for this study is skewed, mean as a measure of central tendency may be influenced by outliers which may not represent the actual data as a whole.
Median	The median is the measure of central tendency of the age of the colorectal cancer cases in which the age is arranged from smallest to largest value. If there is an odd number colorectal cancer cases, the median age is the middle value. If there is even number of colorectal cancer cases, the median age is the average of the age of the two middle cases. Median is a better measure of central tendency for this study as median represents the whole data and is less affected by the skewed data or outliers.

Percentage	The percentage is an expression of proportion of colorectal cancer cases multiplied by a hundred. It is also used to depict each variable (in percentage) in this study pertaining to demographics, clinical and treatment characteristics.
Standard deviation	This is a measure of dispersion of the age of the colorectal cancer cases relative to the mean age. Standard deviation is calculated by subtracting the mean from each age then squaring the difference. Subsequently, the differences are summed then divided by n minus 1 then the square root is taken.