NOTIFIABLE MEDICAL CONDITIONS SURVEILLANCE SYSTEM

the National Institute for Communicable Diseases

# 1 Introduction

R markdown is a simple and easy to use plain text language used to combine your R code, results from your data analysis (including plots and tables) and written commentary into a single nicely formatted and reproducible document (like a report, publication, thesis chapter or a web page like this one). There is also quarto which is a next generation Rmarkdown which can handle multiple languages and is slightly faster, however it does not handle some outputs to word as well Rmarkdown (like figure and table cross referencing) although it will likely handle these in future updates.

# Purpose

The purpose of this session (and document) is to show you how to produce some publication ready tables and publish it to a word document (.docx) from Rmarkdown. We will mainly be using tools from the [tidyverse](https://www.tidyverse.org/packages/), [gtsummary](https://www.danieldsjoberg.com/gtsummary/), [flextable](https://ardata-fr.github.io/flextable-book/), and [ggplot2](https://ggplot2.tidyverse.org/) packages.

# 2 Methods

A selection of code chunks will be given with built in data sets to run. Then a dataset will be given to import and and code chunks to run that are similar to the Monthly [NMCSS Surveillance](https://www.nicd.ac.za/nmc-overview/nmc-monthly-surveillance-report/) report.

Table 2.1: The first summary table caption

| **Characteristic** | **N = 200**1 |
| --- | --- |
| Chemotherapy Treatment |  |
| Drug A | 98 (49%) |
| Drug B | 102 (51%) |
| Age | 47 (38, 57) |
| Unknown | 11 |
| Marker Level (ng/mL) | 0.64 (0.22, 1.39) |
| Unknown | 10 |
| T Stage |  |
| T1 | 53 (27%) |
| T2 | 54 (27%) |
| T3 | 43 (22%) |
| T4 | 50 (25%) |
| Grade |  |
| I | 68 (34%) |
| II | 68 (34%) |
| III | 64 (32%) |
| Tumor Response | 61 (32%) |
| Unknown | 7 |
| Patient Died | 112 (56%) |
| Months to Death/Censor | 22.4 (16.0, 24.0) |
| 1n (%); Median (IQR) | |

See the Table 2.1 for the first output. We can also get values from the table, for instance, the median(IQR) age is 47 (38, 57).

Table 2.2: The second summary table caption

| **Characteristic** | **Drug A**, N = 981 | **Drug B**, N = 1021 |
| --- | --- | --- |
| Age | 46 (37, 59) | 48 (39, 56) |
| Unknown | 7 | 4 |
| Marker Level (ng/mL) | 0.84 (0.24, 1.57) | 0.52 (0.19, 1.20) |
| Unknown | 6 | 4 |
| T Stage |  |  |
| T1 | 28 (29%) | 25 (25%) |
| T2 | 25 (26%) | 29 (28%) |
| T3 | 22 (22%) | 21 (21%) |
| T4 | 23 (23%) | 27 (26%) |
| Grade |  |  |
| I | 35 (36%) | 33 (32%) |
| II | 32 (33%) | 36 (35%) |
| III | 31 (32%) | 33 (32%) |
| Tumor Response | 28 (29%) | 33 (34%) |
| Unknown | 3 | 4 |
| Patient Died | 52 (53%) | 60 (59%) |
| Months to Death/Censor | 23.5 (17.4, 24.0) | 21.2 (14.6, 24.0) |
| 1Median (IQR); n (%) | | |

See 2.2to see the second output of the session.

Table 2.3: The second summary table caption

| **Characteristic** | **Overall**, N = 2001 | **95% CI**2 | **Drug A**, N = 981 | **95% CI**2 | **Drug B**, N = 1021 | **95% CI**2 | **p-value**3 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Age | 47 (38, 57) |  | 46 (37, 59) | 44, 50 | 48 (39, 56) | 45, 50 | 0.7 |
| Unknown | 11 |  | 7 |  | 4 |  |  |
| Marker Level (ng/mL) | 0.64 (0.22, 1.39) |  | 0.84 (0.24, 1.57) | 0.83, 1.2 | 0.52 (0.19, 1.20) | 0.65, 0.99 | 0.085 |
| Unknown | 10 |  | 6 |  | 4 |  |  |
| T Stage |  |  |  |  |  |  | 0.9 |
| T1 | 53 (27%) | 21%, 33% | 28 (29%) | 20%, 39% | 25 (25%) | 17%, 34% |  |
| T2 | 54 (27%) | 21%, 34% | 25 (26%) | 17%, 35% | 29 (28%) | 20%, 38% |  |
| T3 | 43 (22%) | 16%, 28% | 22 (22%) | 15%, 32% | 21 (21%) | 13%, 30% |  |
| T4 | 50 (25%) | 19%, 32% | 23 (23%) | 16%, 33% | 27 (26%) | 18%, 36% |  |
| Grade |  |  |  |  |  |  | 0.9 |
| I | 68 (34%) | 28%, 41% | 35 (36%) | 26%, 46% | 33 (32%) | 24%, 42% |  |
| II | 68 (34%) | 28%, 41% | 32 (33%) | 24%, 43% | 36 (35%) | 26%, 45% |  |
| III | 64 (32%) | 26%, 39% | 31 (32%) | 23%, 42% | 33 (32%) | 24%, 42% |  |
| Tumor Response | 61 (32%) | 25%, 39% | 28 (29%) | 21%, 40% | 33 (34%) | 25%, 44% | 0.5 |
| Unknown | 7 |  | 3 |  | 4 |  |  |
| Patient Died | 112 (56%) | 49%, 63% | 52 (53%) | 43%, 63% | 60 (59%) | 49%, 68% | 0.4 |
| Months to Death/Censor | 22.4 (16.0, 24.0) |  | 23.5 (17.4, 24.0) | 19, 21 | 21.2 (14.6, 24.0) | 18, 20 | 0.14 |
| 1Median (IQR); n (%) | | | | | | | |
| 2CI = Confidence Interval | | | | | | | |
| 3Wilcoxon rank sum test; Pearson's Chi-squared test | | | | | | | |

Table 2.4: The regression summary table caption

| **Characteristic** | **exp(Beta)** | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| Age | 1.00 | 1.00, 1.01 | 0.10 |
| Marker Level (ng/mL) | 0.99 | 0.91, 1.08 | 0.9 |
| T Stage |  |  |  |
| T1 | — | — |  |
| T2 | 1.06 | 0.88, 1.28 | 0.5 |
| T3 | 1.11 | 0.90, 1.37 | 0.3 |
| T4 | 1.47 | 1.21, 1.79 | <0.001 |
| Grade |  |  |  |
| I | — | — |  |
| II | 1.03 | 0.86, 1.23 | 0.7 |
| III | 1.18 | 1.00, 1.40 | 0.053 |
| Tumor Response | 0.77 | 0.66, 0.90 | 0.001 |
| 1CI = Confidence Interval | | | |

We can also make a plot:

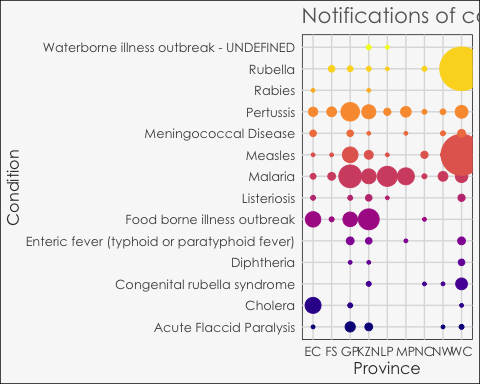


Table 2.5: The number of notifications that are suspected and confirmed for category 1 conditions.

| **Condition** | **Confirmed**,  N = 4151 | **Suspected**,  N = 6461 | **Overall**,  N = 10611 |
| --- | --- | --- | --- |
| Acute Flaccid Paralysis | 0 | 21 | 21 |
| Cholera | 1 | 31 | 32 |
| Congenital rubella syndrome | 19 | 0 | 19 |
| Diphtheria | 0 | 6 | 6 |
| Enteric fever (typhoid or paratyphoid fever) | 11 | 6 | 17 |
| Food borne illness outbreak | 0 | 105 | 105 |
| Listeriosis | 10 | 4 | 14 |
| Malaria | 159 | 45 | 204 |
| Measles | 24 | 229 | 253 |
| Meningococcal Disease | 0 | 18 | 18 |
| Pertussis | 58 | 63 | 121 |
| Rabies | 1 | 1 | 2 |
| Rubella | 132 | 114 | 246 |
| Waterborne illness outbreak - UNDEFINED | 0 | 3 | 3 |
| 1Suspected and confirmed cases are independent and are not totalled - suspected and confirmed cases are distinct. | | | |