NIHR Complex Review Support Unit (CRSU)
19th November 2020

Formatting Data for MetaInsight COVID-19

Introduction

MetaInsight COVID-19 was updated every week up till Monday 19th October 2020; the end of the 5 month period for this pilot app. However, the app now includes the option of uploading your own updated version of the COVID-19 data.

As long as any additional data follows the same format as the data currently extracted and used by the app, all should run smoothly. However, to avoid any trivial issues/errors, this document gives detailed guidance for formatting your own dataset to ensure compatability.

Outcome Sheets

For each outcome, there needs to be two separate sheets: (i) '**COVID**_Outcome_**data**' which contains the meta-analysis data including study characteristics, and (ii) '**COVID**_Outcome_**label**' which contains the label codes for the treatments.
The app is currently fixed to only read the datasheets if they are titled 'COVIDMortalitydata', 'COVIDMortalitylabel', 'COVIDviralconversiondata', 'COVIDviralconversionlabel', 'COVIDSAEdata', and 'COVIDSAElabel'.

Data Sheets

Figure 1 is a screenshot of how the data looks for the sheets containing the meta-analysis data and study characteristics. The first five columns have to be in a specific order and labelled accordingly:

- The **first** column, labelled **StudyID**, should contain the study identifier, starting from 1, then 2, 3, 4… etc. Even if studies span multiple outcomes, within each outcome the study identifier has to be sequential (i.e. studies may have different identifiers across outcomes).
- The **second** column, labelled **Study**, can contain the name (e.g. author, year) of the study.
- The **third** column, labelled **T**, should contain the numerical code for the treatment used in each arm of the study. Labels should be sequential, and the reference treatment (currently 'standard care') needs to be labelled as 1.
- The **fourth** column, labelled **R**, should contain the number of participants with the outcome of interest in each arm of the study.
- The **fifth** column, labelled **N**, should contain the total number of participants in each arm of the study.

The remaining columns, the study characteristics, do not have any specific formatting rules, for the data itself or the column headers. However, the app is fixed to read them in the order listed below and relies on all columns being present. Any additional columns will be ignored by the app. To ensure the filtering option in the app is effective, try to format phrases consistently (e.g. 'Moderate' and 'moderate' will be treated separately by the filter option).

- **Follow-up days**, the number of days participants were followed up for during the study.
- **Dose**, the treatment dosage given in each arm of the study. Currently, a forward slash (/) is synonymous for 'not applicable', and is used for treatments such as 'standard care'.
- **Duration**, the duration of the treatment given in each arm of the study. Currently, a forward slash (/) is synonymous for 'not applicable', and is used for treatments such as 'standard care'.

- **Risk of bias**, the overall risk of bias for the study. Currently, the options are 'low', 'some', and 'high', and were extracted from the covid-nma.com assessment of risk of bias. Within a study, risk of bias may differ between outcomes.
- **Patient characteristics**, the severity of COVID-19 symptoms being presented by participants in the study. Currently, the options are combinations of 'mild', 'moderate', 'severe', and 'critical', and were extracted from the covid-nma.com assessment of the study.
- **Country**, the country (or countries) where the study was carried out.
- **Time of outcome measure**, the number of days from the start of the study at which the outcome of interest was measured at (e.g. mortality 14 days from treatment commencement). Currently, if the study reported results at multiple time points, the result at the later time was taken.
- **Included**, whether or not the study is currently connected to the primary network when on default settings. This column is for descriptive purposes only; the app will automatically include and analyse studies connected to the primary network when on default settings. The current options are 'Yes' and 'Not yet', to indicate that unconnected studies may become connected as more studies are added to the evidence base.

StudyID	Study	T	R	N	Follow_up_days	Dose	Duration	Risk_of_bias	Patient_characteristics	Country	Outcome_days	Included
1	CaoB	1	25	100	28	/	/	some	severe	China	14-28	Yes
1	CaoB	2	19	99	28	400mg+100mg	14 days	some	severe	China	14-28	Yes
2	Yueping	1	0	17	21	/	/	some	mild/moderate	China	14-28	Yes
2	Yueping	2	0	34	21	400mg+100mg	7-14 days	some	mild/moderate	China	14-28	Yes
2	Yueping	4	0	35	21	200mg	7-14 days	some	mild/moderate	China	14-28	Yes
3	ChenJ_Zheji	1	0	15	14	/	/	high	moderate	China	14-28	Yes
3	ChenJ_Zheji	3	0	15	14	400mg	5 days	high	moderate	China	14-28	Yes
4	ChenC	4	0	120	7	200mg	7 days	some	moderate to critical	China	7	Yes
4	ChenC	5	0	120	7	1600mg initial, 60	7 days	some	moderate to critical	China	7	Yes
5	Zhong	6	7	9	30	/	7 days	some	severe/critical	China	14-28	Yes
5	Zhong	7	3	8	30	1200mg	7 days	some	severe/critical	China	14-28	Yes
6	WangY	6	10	78	28	/	10 days	low	severe	China	14-28	Yes
6	WangY	8	22	158	28	200mg initial 100	10 days	low	severe	China	14-28	Yes
7	Hung	9	0	86	30	400mg/100mg + 4	14 days	some	mild/moderate	Hong Kong, China	14-28	Yes
7	Hung	2	0	41	30	400mg/100mg	14 days	some	mild/moderate	Hong Kong, China	14-28	Yes
8	LouY	5	0	10	14	1600 or 2200mg i	up to 14 days	some	unclear severity	China	14-28	Yes
8	LouY	10	0	10	14	80mg	days 1, 4 and 7	some	unclear severity	China	14-28	Yes
8	LouY	11	0	10	14	400mg/100mg or	MISSING	some	unclear severity	China	14-28	Yes

Figure 1 Screenshot of datasheet for main meta-analysis data

Label sheets

Figure 2 is a screenshot of how the data looks for the sheets containing the label codes for treatments. It contains two columns that need to be formatted and labelled as described below.

- Column for numerical treatment codes, labelled **Number**, should contain a unique number for each treatment, starting from 1, then 2, 3, 4… etc. The reference treatment (currently 'standard care') needs to have the numerical code of 1. Even if treatments span multiple outcomes, within each outcome the numerical codes have to be sequential (i.e. treatments may have different numerical codes across outcomes).
- Column for the treatment name, labelled **Label**, should contain the name of the treatment for the corresponding numerical treatment code. Characters for the names are limited to alphanumeric characters and underscores (_).

Number	Label
1	Standard_care
2	Lopinavir_Ritonavir
3	Hydroxychloroquine
4	Umifenovir
5	Favipiravir
6	Placebo
7	a_Lipoic_acid
8	Remdesivir_10days
9	LopRitRibIFNb1b
10	Baloxavir_marboxil

Figure 2 Screenshot of datasheet for treatment code labels