Are the ISO country codes redundant in the 2014 WHO Global Tuberculosis Report Dataset?

Yes, the ISO codes in the 2014 WHO Global Tuberculosis Report Dataset are fully redundant with the country name. Both ISO codes can be dropped from the dataset without losing any information.

Dataset

The dataset used is the 2014 WHO Global Tuberculosis Report Dataset which includes data of the number of cases for 56 tuberculosis variants since 1980 for 219 countries.

Methodology

Given the dataset if we select only the country and ISO codes variables and then proceed to group by country we will get one distinct pattern if the data is redundant and multiple if it is not.

```
#Command used to retrieve distinct events in the dataset

select(dataset, country, iso2, iso3) %>%
 distinct() %>%
 group_by(country) %>%
 filter(n() > 1)
```

Results

After applying the tidyverse distinct function we do not get any distinct patterns which means the ISO codes are completely redundant with the country name, as expected.

Conclusion

The ISO codes are redundant and thus can be dropped without losing any information. The people who made the data set made no mistakes when inputing the ISO codes.

Wickham H (2022). World Health Organization TB data. R package version 1.0.2, https://tidyr.tidyverse.org/reference/who.html

```
"Written by Carlos Vintimilla"
```