

02 Selection

35762384

17/02/2021

Introduction The availability of health facility data, such as location, capacity and resources present, is an important factor needed for decision-making processes, especially in the ongoing COVID-19 pandemic. Examples of uses include planning of interventions, disease surveillance, information for insurance companies and health management information systems (HMIS) (WHO, 2018). It is also often used for research purposes, with many studies utilising facility data to determine accessibility to healthcare or travel times (Hulland et al., 2019). However, availability is often not the issue but the fact that there are multiple sources and the discrepancies between them (Makinde et al., 2018). It is common for different organisations, e.g. non-governmental organisations (NGOs), government departments and other non-profit organisations, to produce their own lists and a need to further investigate differences is noted in studies using this data (Hulland, 2020). The aim of this paper is to do address this point; to analyse and highlight the differences between sources of facility data using Malawi as a case study and mark areas for improvement in the quality of the data.

One source for Malawi is a list compiled by the Kenyan Wellcome Trust Research Programme (KWTRP) for 50 sub-Saharan African countries and was released in 2019 (Maina et al., 2019). It focused on public health facilities, those run by the government, faith-based organisations (FBOs) and NGOs and also removed facilities that only provide specialised care such as psychiatry. It's availability as open-access data as well as the thorough cleaning and validation processes implemented, meant it is often cited in other studies (Falchetta et al., 2020) (Judson et al., 2020) (Dowhaniuk, 2021) (Wariri et al., 2021). Several sources of information for facilities were combined to produce one list and in the case of Malawi, personal communication with health related organisations, data from The Humanitarian Data Exchange (HDX) and The Christian Health Association of Malawi (CHAM) were used (Maina et al., 2019). An important note is that the years of when the information was acquired from these sources vary, with the most recent being 2017 and the personal communication being conducted in 2013. The list is now being hosted by the World Health Organisation (WHO) Global Malaria Programme with the aim to update (WHO, 2019). However, since the publication of this study, there have been no updates or changes. Therefore, there is a need to compare with other sources, especially as the age of this data might mean new developments are not captured which could affect the study's popularity.

A newer source is the Master Facility List (MFL) from and managed by the Malawi Ministry of Health (MOH, 2021). The WHO recommends that every country produces a MFL with the aim of it being the primary source and describes how to develop it in (WHO, 2019). It must be accessible, regularly updated and validated. Although many African countries have made steps towards formulating a MFL under these guidelines, issues of access and missing elements are often encountered, making its usage more difficult (South et al., 2021). With Malawi, the MFL is openly available. However, its validation methods are not made clear and information on the most recent update is only available when selecting a specific facility and is not part of the dataset that can be downloaded. Despite this, the fact that it is maintained by the Ministry of Health and so has the potential to be incorporated into the health system and that it is a more recent source, it is worth investigating how this list compares to others.

Other sources are mostly dependent on information contributed by volunteers, of which healthsites.io is a global project aiming to map every health facility running (Healthsites, 2021). It works with OpenStreetMap, which provides the baseline map as well as the methods to input data, and since its establishment in 2016

has recorded over 900,000 facilities. Anyone can contribute and effort has been put into validation processes, which includes a Location Validation Index, a score that reflects other users verifying that facility exists. The data is freely available and access to the most recent version can be gained through several formats such as an API. Other lists can also be incorporated into healthsites.io and there is a process outlined for the import of facility lists from national ministries of health. However, similarly with other lists, it struggles with completeness and it has been shown that less than 2% of healthsites.io data for sub-Saharan Africa contains attributes describing capacity (South et al., 2021). It seems that the quality or functionality of these lists is limited across all of these sources.

Quality data is important but this is often neglected or not investigated. There are several issues that are prominent with facility data from sub-Saharan Africa. Lists are often missing key elements such as capacity, equipment and services they provide. Not only this, problems with missing coordinates can also occur, with the WHO-KWTRP data reporting 9 missing coordinates for Malawi (Maina et al., 2019). These issues are present but there is not much research highlighting this, especially when comparing between sources, but this is important for studies that go on to utilise these lists. The outbreak of COVID-19 brought a greater emphasis on the need for this. Several countries have allowed open access to their facility data, encouraging external research to aid the response to the pandemic and improvement of the data itself. For example, research investigating the ability of health facilities to increase capacity and the identification of people vulnerable due to various factors in Kenya were performed with open facility data (Barasa et al., 2020) (Macharia et al., 2020). Therefore, maintenance and quality control of facility lists are essential in providing accurate data and contributing to valid research.

The aim for this paper is to provide a reproducible summary of facility data for Malawi, discuss differences that are relevant to potential stakeholders and make apparent areas for improvement. The following hypotheses have been developed to address this:

1. The same number of health facility locations are recorded for Malawi across healthsites.io, WHO-KWTRP and their MFL
2. The health facility locations and proportion of facility types recorded in the Malawi MFL are similar to those being used in current global analyses
3. Hospital locations stored in healthsites.io are currently not a good representation of those available from the Malawi MFL

Methods

- Websites from where I downloaded
- What was done to the data pre-analysis, e.g. removing NAs
- Software used (packages?)

Results The data downloaded contained 1546 facilities in the MFL, 648 in the WHO-KWTRP list for Malawi and 248 in healthsites.io. After removal of missing coordinates and duplicates, there are 1424, 638 and 234 facilities in the MFL, WHO-KWTRP and healthsites.io data respectively (figure 1). The MFL contains more than 6 times the number of facilities compared to healthsites.io.

Both the MFL and WHO-KWTRP contain some missing coordinates. Upon inspection, there are 119 missing values and 62 coordinates that are not in Malawi, which were either inputted incorrectly or not known, in the MFL. In the WHO-KWTRP, there are 9 missing coordinates and this was stated in (Maina et al., 2019). Healthsites.io carries some missing attributes but all coordinates are present and reside within Malawi. Analysis into duplicates within lists shows that names for 5 facilities in the MFL were duplicated at least once and these also have similar coordinates up to at least 2 decimal places. Of these 5 facilities, 3 share the same entries in the other attribute columns while the remaining 2 differ in type. In the WHO-KWTRP, 1 health centre appears twice with the same attributes and coordinates. With healthsites.io, there are 18

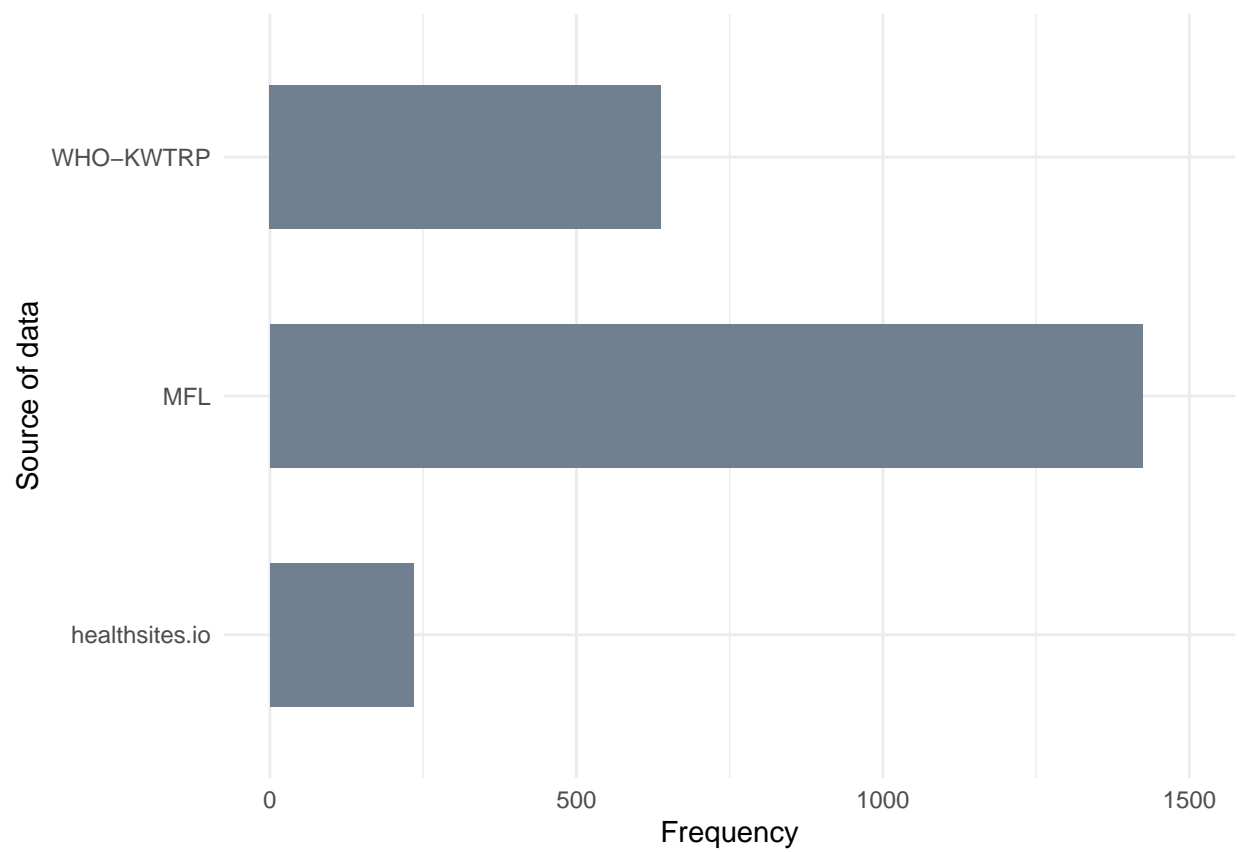


Figure 1: Total number of facilities recorded in each data source.

names that were inputted more than once and all apart from 2 have similar coordinates up to at least 2 decimal places. Searching for duplicate coordinates rather than names also gave matches. In healthsites.io, 4 facilities that were also part of the previous 18 have identical coordinates. The time stamp on these facilities indicate that their duplicates were inputted at the exact same time. The WHO-KWTRP list does not have duplicates while the MFL returned 44 distinct coordinates that were repeated at least more than once. One of these cases is repetition of one location for 24 facilities, of different types and names, in Blantyre. However, some results are due to no coordinates being available and for example, (-1,1) was recorded instead.

Figure 1 shows method and number of duplicate data points that have been removed as part of the analysis. As discussed, duplicates are considered by name or coordinates. If the coordinates and main attributes match, the duplicate is removed. Therefore, majority of facilities with duplicate coordinates in the MFL were not removed due to possessing different names or were of different types. Any missing coordinates were also omitted before further analysis. The 62 incorrect coordinates in the MFL were kept and not included in this.

031, -16.73075	2	Same attributes. Coordinates are identical
712, -14.65773	2	Same attributes. Coordinates identical
2, -1	2	Same attributes except 'Common name'
3, -2	2	Same attributes except 'Common name'
223, -9.67965073595659	2	Same type and name. Coordinates identical
404, -13.9445974818355	2	Same type, name and time inputted. Coordinates identical
144, -13.9878648128351	2	Same type and name. Coordinates identical
697, -13.9903576746579	2	Same type, name and time inputted. Coordinates identical
211, -14.0263392571441	2	Same type, name and time inputted. Coordinates identical
588, -14.2017512005316	2	Same type and name. Coordinates identical
1524, -13.987080182879	2	Same type, name and time inputted. Coordinates identical
1233, -9.6695534762986	2	Same type and name. Coordinates identical
225, -15.2615526872826	2	Same type, name and time inputted. Coordinates identical
087, -14.0066068321501	2	Same type, name and time inputted. Coordinates identical
042, -15.3855159746797	2	Same type and name. Coordinates identical
846, -16.2777391528508	2	Same type and name. Coordinates identical

Figure 2: Table of duplicate facilities removed from each source. 'No.of.inputs' refers to number of times that facility was duplicated and rationale is reported under 'reason'.

Hypothesis 2 With the WHO-KWTRP list often used in research, a comparison to the MFL is made here. Setting aside the groups of unclassified and private in the MFL, both the MFL and WHO-KWTRP have similar numbers of classification types (figure 2). However, there is a difference in the amount of specificity. The WHO-KWTRP provides 5 categories for hospitals while the MFL only provides 3. In contrast, the MFL separates health posts and dispensaries while WHO-KWTRP combines these facilities. The private and unclassified groups not present in WHO-KWTRP only make up 0.4% of the number of facilities in the MFL. Much of the variation between these two sources is accounted for by the number of each type

of facilities. Hospitals constitute 8% of the total in the MFL and 13% in WHO-KWTRP. Health centres form most of the facilities in the WHO-KWTRP list, 71%, while the closely related clinics only form 3%. In the MFL, there is more of an even spread of clinics and health centres, which constitute 39% and 35% respectively.

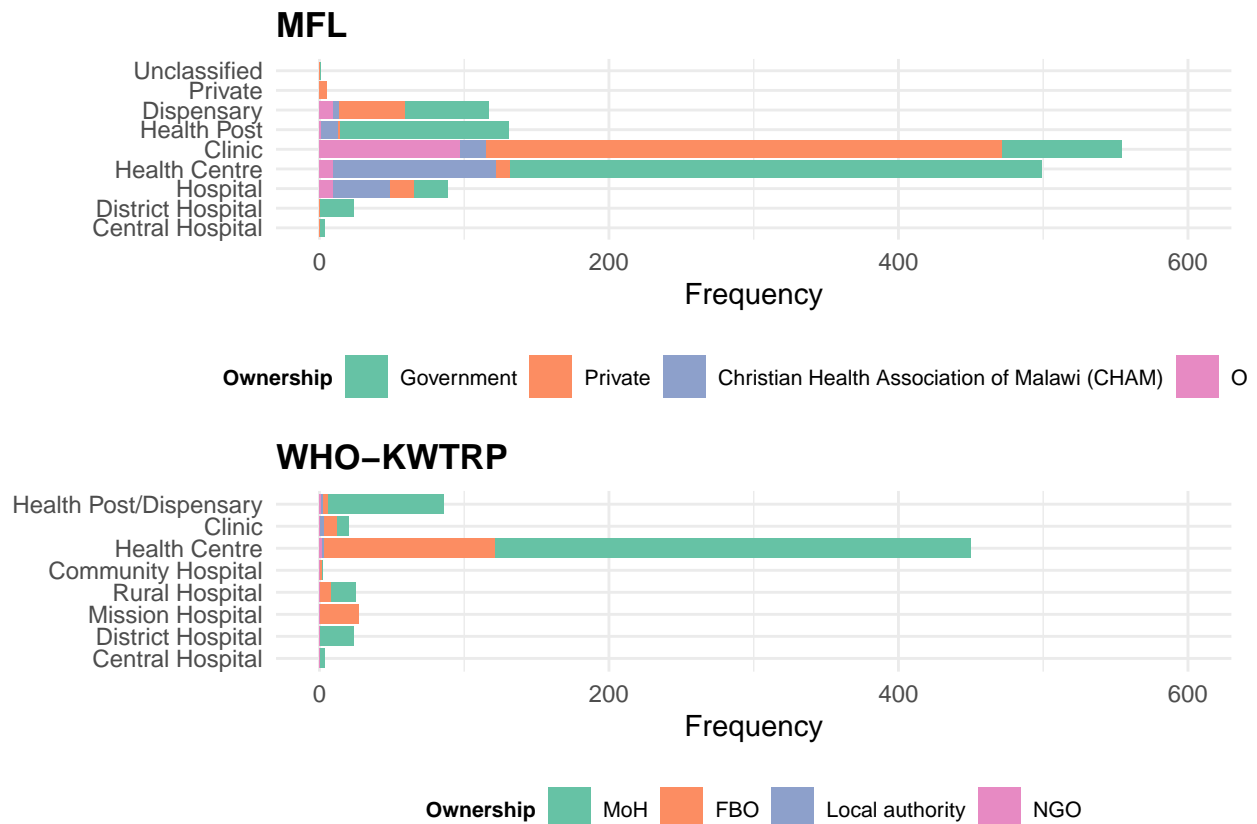


Figure 3: Number of each type of facility in the MFL (top) and WHO-KWTRP (bottom), with proportions corresponding to ownership for each type coloured.

Figure 3 shows the point locations of facilities from both sources and an interactive version with all facility types can be viewed online. By observing this map, it appears that the same central hospitals are recorded in both lists and majority of district hospitals, apart from 2, are also matching (figure 3). Distribution of the 3 other hospital categories in WHO-KWTRP is predominantly similar to that of ‘hospital’ in the MFL, with the exception of some central and southern districts. It also confirms the MFL has a significant number of clinics that are not accounted for in the WHO-KWTRP list, which only reports a small number in the central and southern regions (figure 4).

Comparison between private facilities is also made here, as the WHO-KWTRP data does not include this. In the MFL, 30% of all facilities are privately owned. Majority consist of 356 clinics, 45 dispensaries and 16 hospitals. Distribution across the country indicates many are present in Blantyre and Lilongwe, with 95 and 67 private facilities respectively (figure 5).

Hypothesis 3 Within healthsites.io, 158 hospitals are recorded which is greater than both the MFL and WHO-KWTRP, with 117 and 82 hospitals respectively (figure 6). Healthsites.io does not provide a breakdown of hospitals into groups. A look into the distribution by district shows majority of the difference in hospital number is concentrated around 3 districts in the south (figure 7). Blantyre district has 39 hospitals in healthsites.io compared to 14 in the MFL and both its neighbouring Mwanza and Thyolo districts have

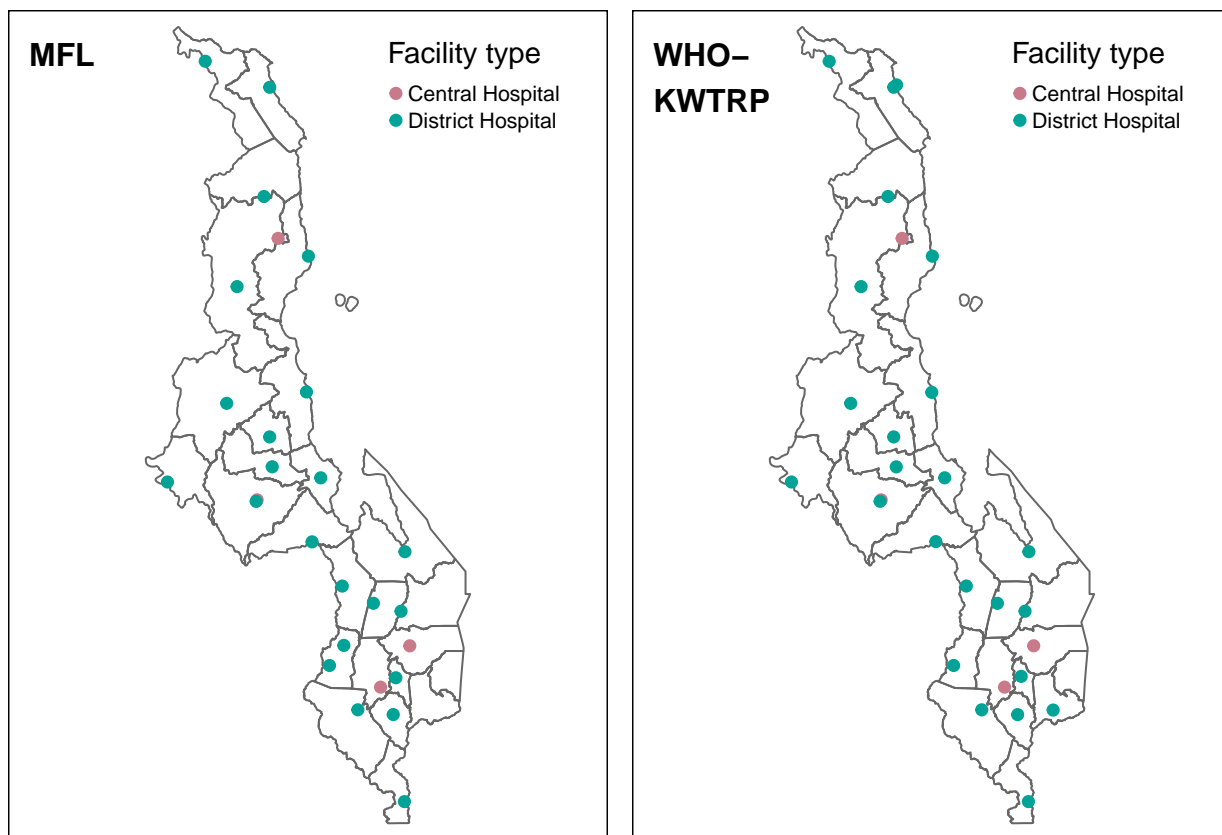


Figure 4: Locations of central and district hospitals in the MFL and WHO-KWTRP.

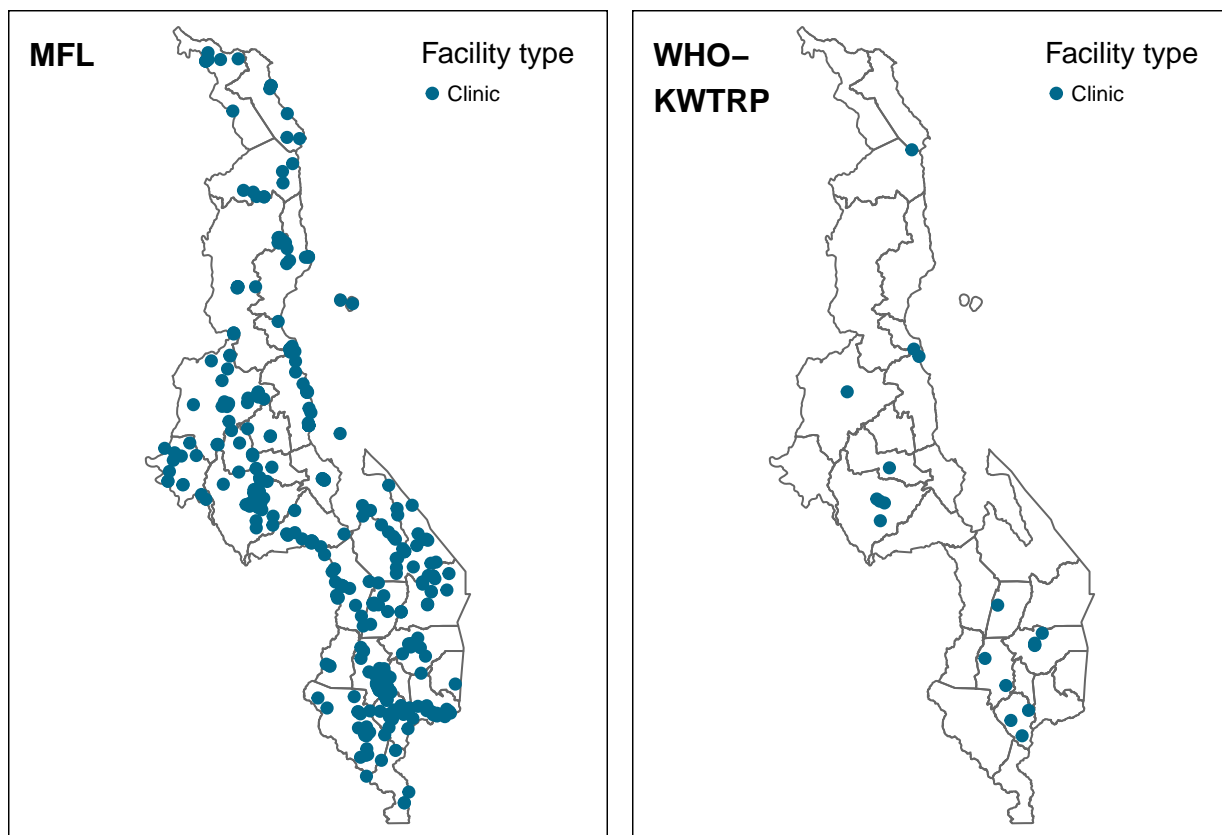


Figure 5: Locations of clinics in the MFL and WHO-KWTRP.

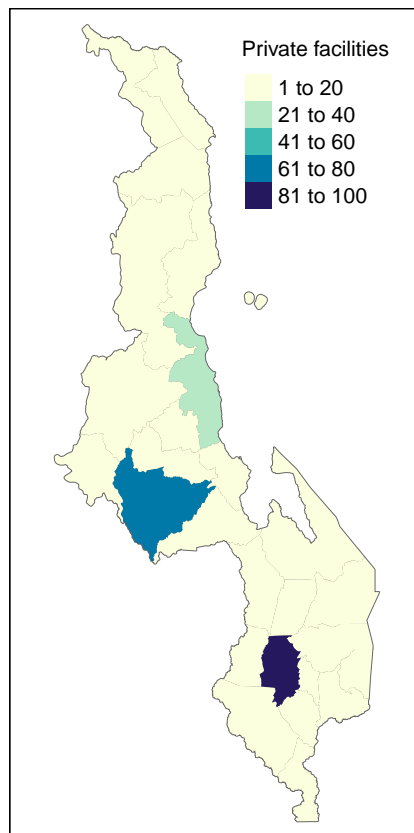


Figure 6: Number of private facilities per district in the MFL.

9 and 7 additional hospitals in healthsites.io respectively. Instances where the MFL has more hospitals in a district, the number does not exceed more than 4 facilities. A note here is that 4 hospitals in the MFL did not have accurate coordinates and so were dropped from the total of 117 in this analysis.

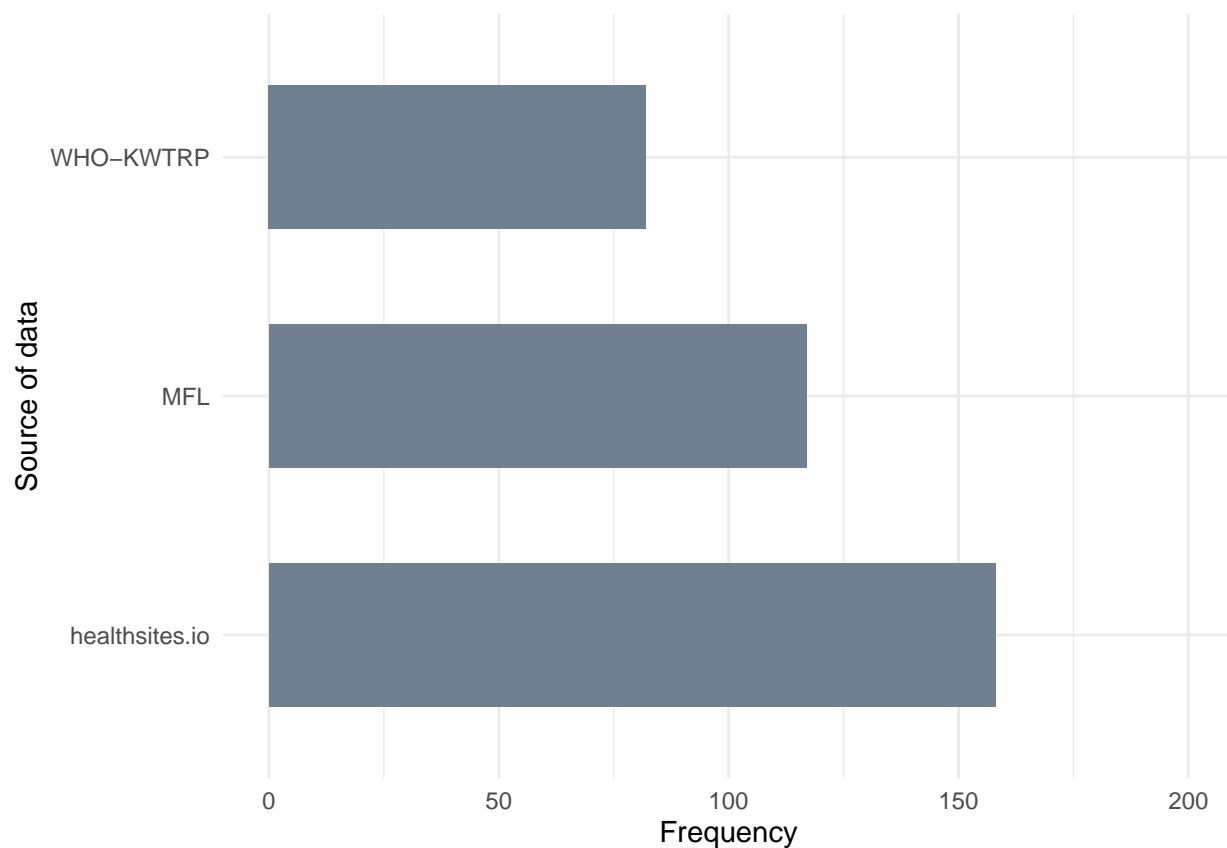


Figure 7: Number of hospitals in each data source.

Blantyre talk about distribution of facilities in MFL and WHO across Blantyre (rural & city differences)
(23 private facilities did not intersect)
talk about private in Blantyre
talk about hospitals in Blantyre

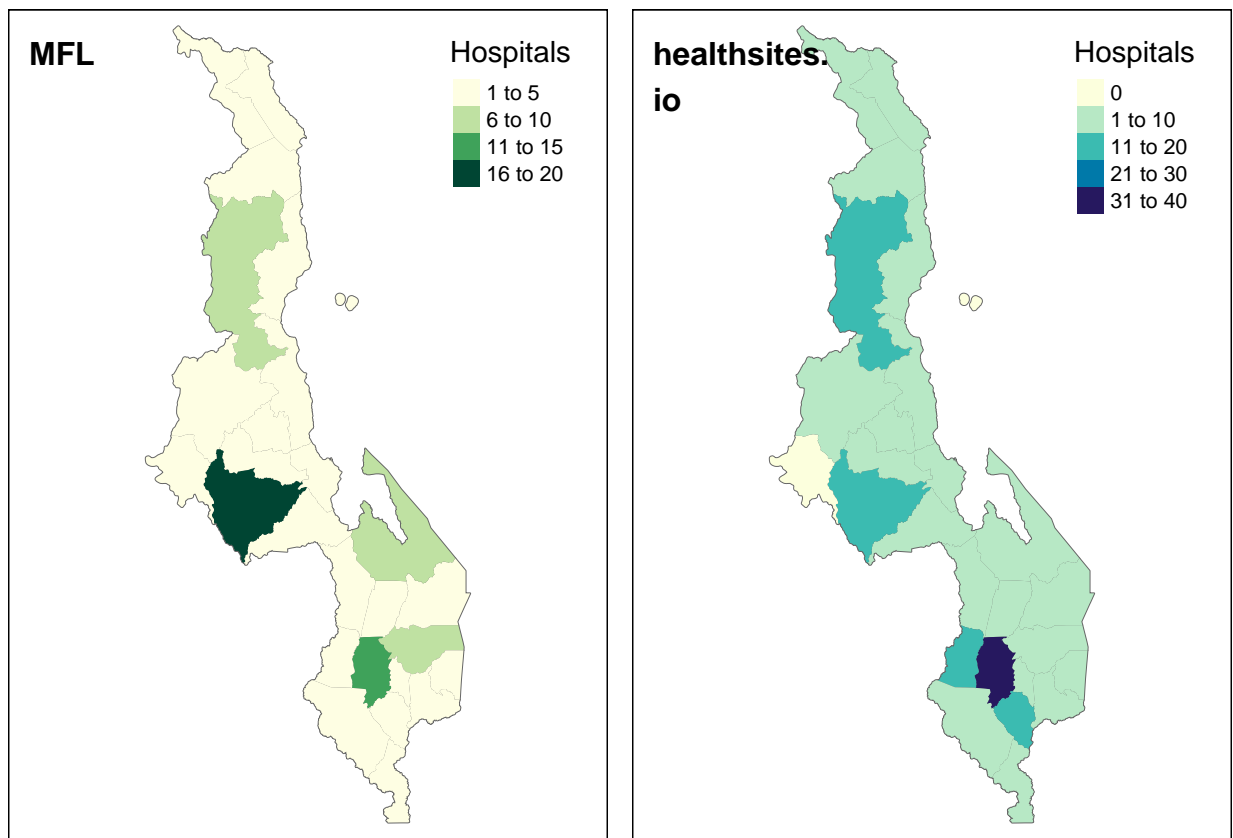


Figure 8: Number of hospitals per district from the MFL (left) and healthsites.io (right).

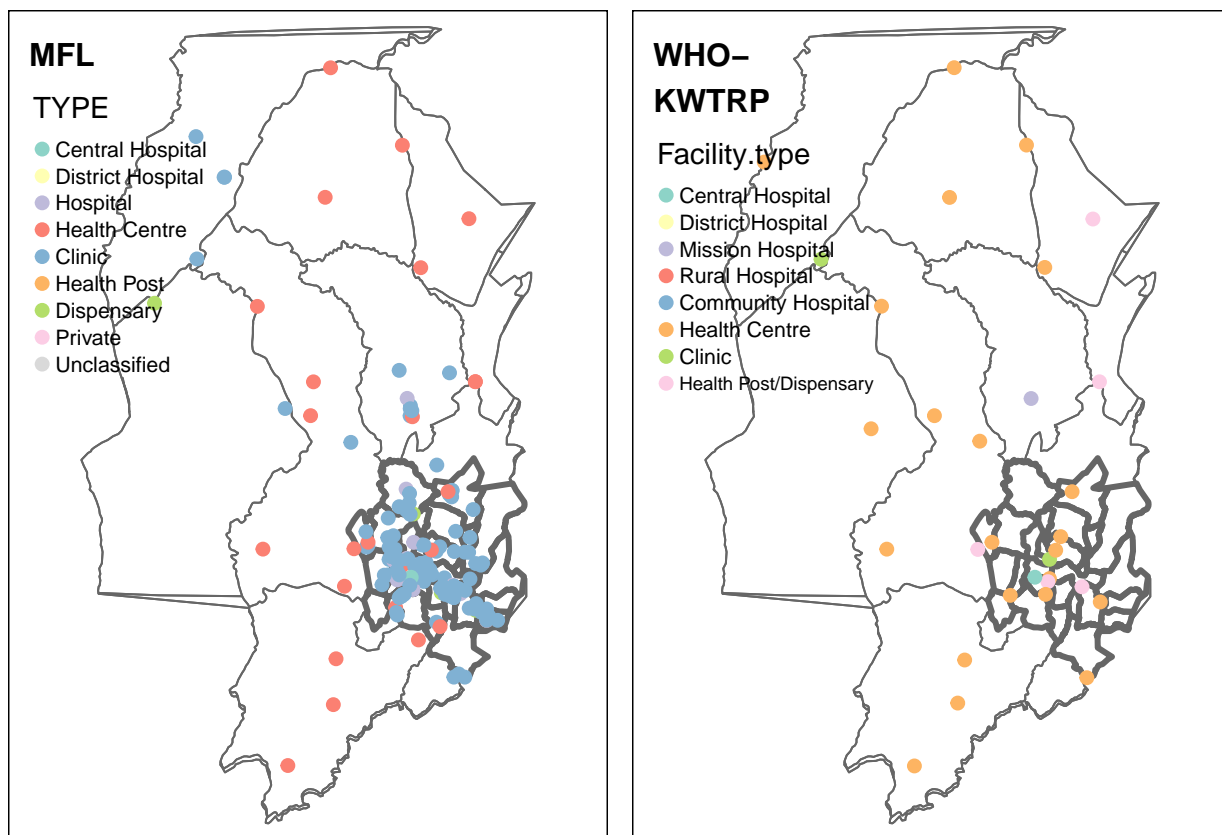


Figure 9: Distribution of facilities present in Blantyre district from both the MFL and WHO-KWTRP.

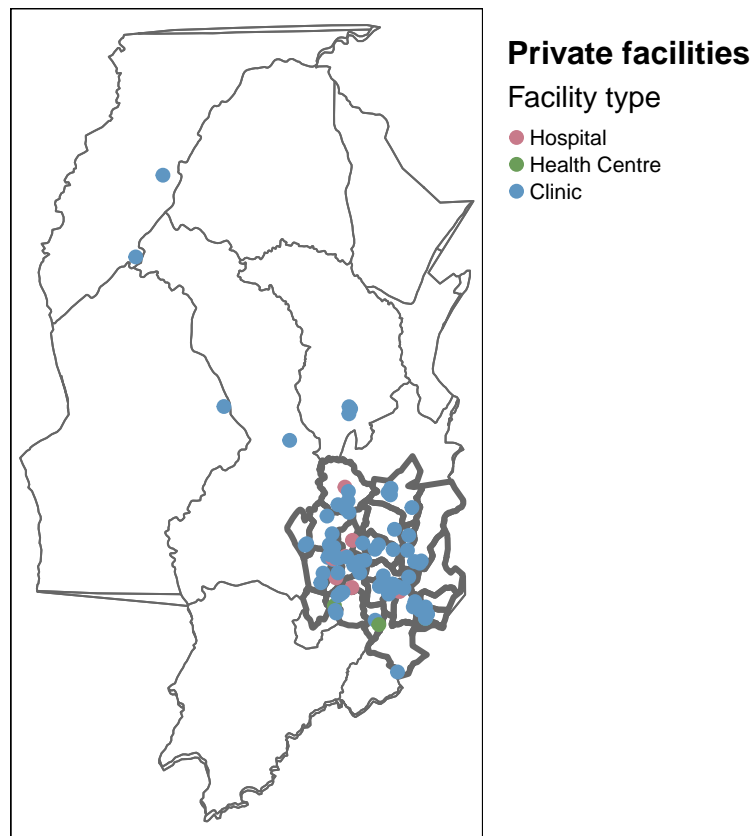


Figure 10: Distribution of private facilities in Blantyre district obtained from the MFL.

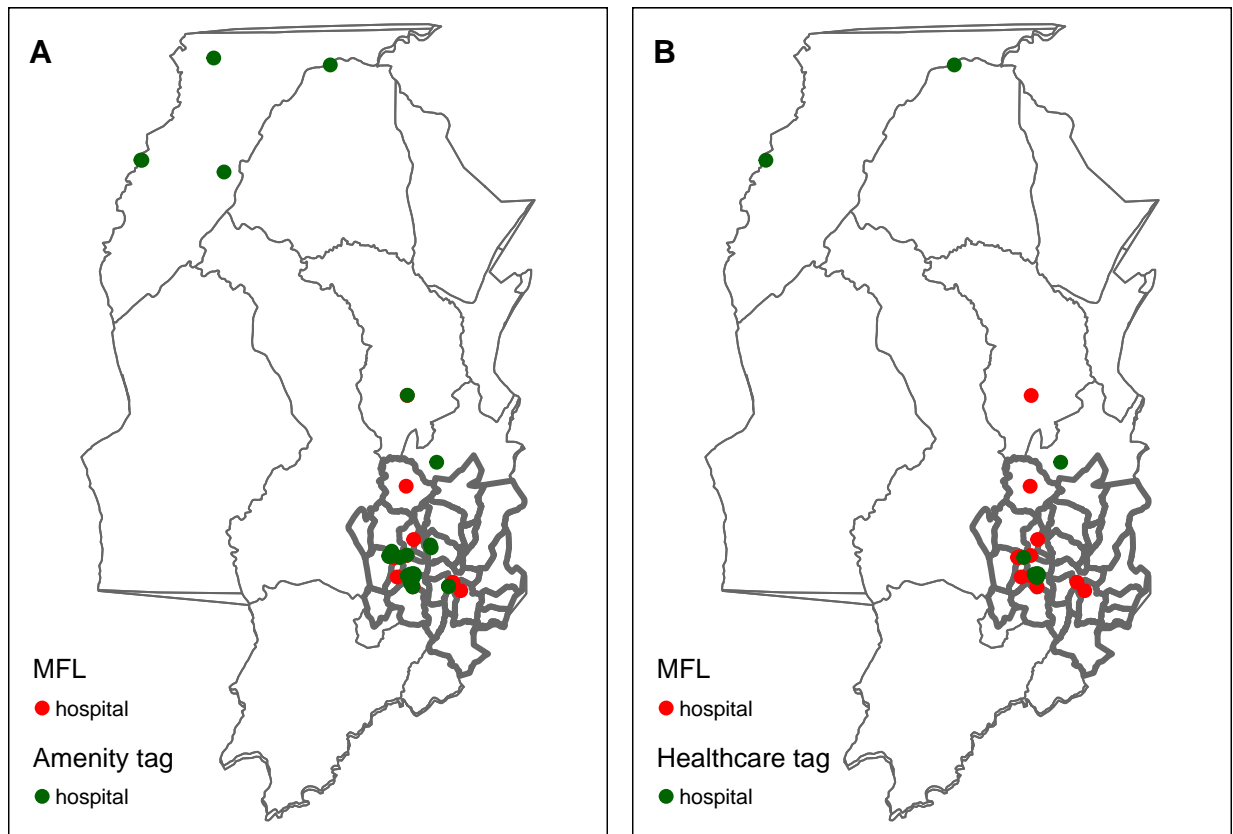


Figure 11: A. Distribution of MFL hospitals and facilities labelled hospital with the amenity tag from healthsites.io in Blantyre district. B. Distribution of MFL hospitals and facilities labelled hospital with the healthcare tag from healthsites.io in Blantyre district.

Discussion This report conveys a number of differences between these facility lists for Malawi and highlights the issues of missing/duplicate values. A small number of facilities were removed from each source as a consequence, which is often not done in research utilising these lists (Hulland et al., 2019). However as mentioned, the search for duplicate coordinates, particularly in the MFL, produced many hits yet these were not removed due to different facilities being accounted for by these locations. Its removal would have larger effects on the proportions of facility types. Although this is subjective and the need for proper validation processes in the MFL is emphasised.

The guidance put together by WHO on the production of a MFL, recommends that district officers or individuals that have access to facilities in their areas be responsible for their validation. This can either be done through visits or telephone calls after which unique identifiers can be assigned (WHO, 2019b). On the Master Health Facility Registry website that hosts the MFL, it mentions that only users with ‘access control rights’ are able to add or edit facilities (MOH, 2021). This indicates there are filters in place that prevent validation from just anyone. However, whether a facility is validated prior to its open release is not mentioned and validation cannot be asserted by viewing a facility online or in the data downloaded. A case in the Philippines, where workshops were organised to bring together Department of Health staff to validate all the facilities in their MFL, shows how a different method might be needed to encourage this process (WHO, 2019b). More than the MFL, healthsites.io also contained many duplicates despite their outline for a validation index (Healthsites, 2021b). Often those facilities were inputted by the same user and at the exact same time, suggesting a possible error in the system. Another issue prominent was misclassification of facilities, in particular they being labelled as hospitals.

On the Wikipedia page of healthsites.io, intended as information for volunteers looking to input facilities, it states that the healthcare and amenity columns include entries on the type of facilities and should be identical (Healthsites, 2021c). However, this was not the case for Malawi and there were a number still labelled as hospital despite their names suggesting otherwise. Instructions on classification are provided on the Wikipedia page, with hospitals relating to facilities offering specialised care and overnight stays while clinics refer to a medical centre with doctors that offer outpatient care (Healthsites, 2021d). In total, healthsites.io gives the option of 5 classifications while the WHO-KWTRP and the MFL data allow for 8 and 9 classes respectively. The WHO-KWTRP data referred to country specific Health Sector Strategic Plans (HSSPs) to define health facilities for each country (Maina et al., 2019). However, it still does not match with the MFL in the case of Malawi, adding another method of classification to consider. These differences make comparison difficult, especially with lower level facilities, when classifications are not uniform and not apparently accessible. One issue is difference in methods, but volunteers tagging a facility incorrectly is also a problem in the case of healthsites.io and further shows the need for validation. A recent study measuring global access to healthcare used data from healthsites.io as well the WHO-KWTRP data and stated only the use of hospitals and clinics (Weiss et al., 2020). Incorrect tagging would have affected the results of this research and potentially inflated the number of hospitals, as is shown in this look into Malawi.

Research using Malawi facility data should also consider the number of private facilities not captured by WHO-KWTRP. Most of these consist of clinics with Blantyre district hosting the largest number. Blantyre is the second largest district in terms of population, around 1 million, of which nearly 60% is estimated to be urban (Mzilahowa et al., 2016). Facilities in rural areas that are not accounted for can impact accessibility measurements and private clinics make up most of the difference between total numbers of WHO-KWTRP and MFL facilities in rural Blantyre. Maina et al., state their reason for exclusion being difficulty in auditing and identifying these facilities. However, 30% is not an insignificant number and stakeholders or researchers need to consider whether to use the WHO-KWTRP data when the additional facilities are not limited to urban areas. Date of data collection is also a potential reason for the discrepancy in total number. As mentioned, the WHO has not updated their list since early 2019 and while MFL updates can be viewed when clicking on specific facilities, it is hard to determine the dates of when facilities have been added to the MFL. A random browse through their site shows the most recent updates span 2019 and early 2020.

- Why is healthsites much less & is it used often in research?
- Distribution of the differences and effect on research

- Attributes -> why they are important and lack of them in lists
- Conclusion