Benchmarking Health Systems in Middle East and North Africa Countries

By Huihui Wang* and Abdo S. Yazbeck

*Corresponding author: hwang7@worldbank.org

There is no conflict of interest that could be perceived to bias this work, acknowledging all financial support and any other personal connections.

Abstract

Health systems are not easy to benchmark, in part because the health sector produces more than one outcome. This paper offers two ways of benchmarking the health systems of countries in the Middle East and North Africa (MENA) focusing on two different outcomes, health status and financial protection. The first approach is by measuring the gap between predicted health outcomes based on country socioeconomic status and actual health outcomes. The second approach is by simply comparing the levels of out-of-pocket (OOP) spending in MENA countries. The paper offers some interesting findings about the large heterogeneity in both health system outcome achievements despite considerable cultural and linguistic similarities in the region. Moreover, three discrete clusters of countries are found on the health status measure. The findings also give specific health system target outcomes for MENA countries to focus their reform efforts.
Motivation

Health systems are not easy to benchmark, in part because the health sector produces more than one outcome. Whether one uses the World Health Organization’s\(^1\) or other frameworks\(^2\) for health sector production, at least three distinct outcomes are associated with any health sector. The most obvious of the three is health status of the population, followed usually with some measures of financial protection for the population from the impoverishing impact of paying out of pocket for care, which is then followed by a third measure related to the satisfaction of the population with the health sector. But even if one looks only at the first outcome, health status, it is not a single measure. There are many different measures of mortality\(^3\) and morbidity including combinations of the two like quality-adjusted life years\(^3\) and disability adjusted life years.\(^4\)

In this paper, we benchmark Middle East and North Africa (MENA) health systems in two ways. The first part is done by using data and methodology from a new global effort\(^5\) updating work from Preston\(^6\) and the World Development Report (WDR) 1993.\(^7\) The new data allow the production of an empirical picture of health system performance of over 160 countries while controlling for important social and economic determinants. In this paper, we also include measures of financial protection as a second dimension of benchmarking for health systems in MENA.

MENA offers a wide array of heterogeneous health systems in countries that share historic, linguistic, and cultural contexts. Table 1 shows some key background information of the MENA countries under analysis. There are significant variations between countries in per capita national income (US$1,300-92,320), life expectancy (62–82 years), percentage of total health
expenditure out of Gross Domestic Product (2–11%), and average years of schooling (4–11 years).

Table 1. Background characteristics of MENA countries

[Insert Table 1]

By looking at performance of the different systems along the two dimensions, health outcomes and financial protection, we allow the identification of success and failures as well as a hard look at potential tradeoffs. Equally important is that such benchmarking helps to set health system expectations and targets for countries to bring in an element of accountability and to influence decision making.

**Health Status**

As noted above, health status can be measured in different ways, typically but not always related to mortality. The most used measure of health status, and likely the most intuitive, is life expectancy at birth (LEB). In 1975, Preston published a paper that looked at the correlation between LEB and per capita national income. That relationship, eventually referred to as the Preston curve, shows a strong relationship but does not assign causality. Even when the Preston curve is calculated in different eras, the relationship remains but the curve shift up presumably due to new medical technologies and public health actions that allow better health outcomes at same national income levels. What the Preston curve also showed is that at any level of per capita national income there are ranges of LEB across countries, in some ways reflecting varied
performance of the health sectors of these countries. In other words, a possible interpretation of the scatter of countries around the curve is in part due to differential performance of their health systems.

We updated the Preston curve calculation here using the most recent available data from World Development Indicators database,\(^8\) LEB (years) and per capita Gross National Income (Atlas method, current US$) in 2014 for all countries in the world. Figure 1 shows that the basic relationship between LEB and per capita national income remains fairly strong as well as the scatter around the regression line. Looking specifically at the MENA countries, a number of things stand out:

1. The LEB and per capita national income relationship is much flatter for the region than globally. Per capita national income does not appear to have as strong a relationship with LEB in the region.

2. The countries with the best performance within their National income range are Lebanon, Morocco, Tunis, Iran, West Bank and Gaza, Jordan, Algeria, and Egypt.

3. Countries that have lower LEB than their National income range are Kuwait, Qatar, Saudi Arabia, the United Arab Emirates, and Iraq.

Figure 1: Updated Preston Curve, 2014

[Insert Figure 1]

In 1993, the World Bank extended the Preston curve analysis in two ways in their flagship publication, the *World Development Report*, which was focused on health. WDR93 noted that
educational stock had been found to also be highly correlated with health status measures like LEB so LEB was regressed against per capita national income and educational stock in a country. WDR93 further used the best regression fit to predict what a country’s LEB should be given national income and educational stock and plotted the difference between predicted and actual LEB. If the predicted LEB is lower than actual LEB then a country is producing a higher level of health status given the development level. If the predicted LEB is lower than actual then the country’s health sector is performing below expectations.

The WDR93 further extended the Preston work by regressing per capita spending on health (both public and private) on level of national income and educational stock and then plotted the difference between predicted and actual health spending. Both predicted versus actual measures were then plotted on a four quadrant graph (cross hairs graph) that highlighted which countries: (q1) spent less than expected on health but produced higher LEB than expected—top left quadrant; (q2) spent less than expected and produced less than expected—bottom left quadrant; (q3) spent more than expected but produced less than expected—bottom right quadrant; and (q4) spent more than expected and produced more than expected—top right quadrant.

We reproduced the WDR93 work with most recent available data among more than 160 countries. Specifically, LEB and per capita spending on health (current US$) in 2014 are regressed against per capita national income (Atlas method, current US$) in 2014 and educational stock\(^b\) as of 2010. We did log transformation for per capita spending on health and per capita national income for the best fit of regression.

We plotted differences between predicted and actual measures in Figure 2, and summarized regression results in Table 2. Figure 2 highlights MENA countries and only shows countries where the gap between expected and actual LEB was between minus 10 and plus 10 years for
presentational purpose. Like the WDR93 effort, 23 years earlier, Figure 2 shows no relationship between per capita spending on health and health status achievement relative to expected. Also like the earlier effort, however, variations in health sector performance in terms of LEB relative to the development level of countries stand out.

Table 2: Summary of regression results

[Insert Table 2]

Figure 2: LEB and per capita health spending relative to per capita national income and educational stock, 2014

[Insert Figure 2]

Figure 2 shows three clusters of MENA countries, worse health status with lower health spending, slightly better health status with slightly lower spending, and better health status with slightly higher spending. When we control for per capita national income and educational stock, Morocco, Lebanon, Algeria, and Tunisia stand out as best performers on LEB. Morocco in 2014 outperformed expected LEB by 7.33 years, Lebanon by 6.16 years, Algeria by 4.8 years, and Tunisia by 4.49 years. What is also interesting is that most Gulf Cooperation Council (GCC) countries are both underspending relative to expectations and underperforming on LEB relative to level of development. In particular, there appears to be high levels of underspending by Qatar, Kuwait, UAE, and Saudi Arabia. Three countries (Iran, Egypt and Bahrain) managed to achieve better health outcomes at a relatively lower spending, implying good efficiency of their health systems in using resources.
The global benchmarking effort\textsuperscript{5} also looked at health system performance over time. Basically the regression based analysis comparing actual and predicted LEB was repeated every 10 years from 1974 to 2014. Unfortunately not every country has the required data at every time period especially for measure of educational stock, but there were enough country points that made it possible to identify countries that consistently perform better than expected, countries that consistently underperform, as well as other patterns. For the Middle East, we found:

1. Countries that consistently beat predicted LEB given National Income and education stock were: Tunisia, Morocco and Egypt (five data points form 1974-2014), Jordan and Bahrain (four data points), and Syria (four points, until 2004 with no data in 2014).

2. On the other extreme, two countries were repeatedly below the predicted LEB, Qatar and UAE, though there were only two data points out of the possible five. .. Saudi Arabia had four points below expected LEB (1974, 1984, 2004, and 2014) and one year above expected LEB (1994).

**Financial Protection**

A second important health system outcome is the extent to which the health sector protects (or fails to protect) people from falling into poverty due to out-of-pocket (OOP) spending on health services. This outcomes can be measured in a number of ways ranging from detailed analysis of the number of people that either fall into poverty—or deepen the level of poverty—due to out of pocket spending on health care, to the much simpler measure of percentage of OOP spending on health care out of total health spending. For a benchmarking exercise, availability of data for a
large number of countries becomes the main challenge when selecting indicators, and this makes OOP spending an excellent candidate for a proxy for financial protection. Figure 3 ranks MENA countries where data was available in 2014 from the countries with the highest to lowest percentage of OOP spending. High OOP would then reflect a higher financial risk to countries than low OOP.

What is quickly obvious from Figure 3, especially when compared with the health status outcomes presented in Figures 1 and 2, is that performance on these two dimensions seems to be reversed. Figure 2 tells us that controlling for level of national income and educational stocks, Morocco is the best performer on LEB and yet Figure 3 shows Morocco with the highest percentage of OOP in the region. Conversely, the GCC countries did not perform well on the LEB relative to national income and educational stock, but lead the region in terms of lowest percentage of OOP spending.

Figure 3: Out-of-pocket spending on health care services as a percentage of total spending in 2014

Another advantage of using OOP spending data is that it is available over a fairly long period of time for most countries allowing tracking changes over time. Figures 4 presents 20 year time series on OOP for health care for five countries. The top and bottom lines are clearly for the country with highest OOP, Morocco, and lowest OOP, Oman. For Morocco, there appears to be a shift to a higher level of OOP in the early 2000s, but stability before and after that. For Oman, there was considerable stability in the numbers for all but the last two years where there is a
pronounced downward shift. A policy question would then be: What changed in the health sector that shift long term patterns of OOP spending.

The other three countries in Figure 4 are all GCC countries. Each of them, however, shows shifts in different directions. The most consistent of the three is Saudi Arabia which has a mostly declining trend with a small uptick in 2007 till 2011 then a return to the main pattern. The other two, UAE and Bahrain, have somewhat opposing patterns. UAE started relatively low then has a sizable increase for a long stretch then return to the previous level from 2009 on. Bahrain, on the other hand, shows a steady decline in OOP then a clear uptick in 2011. What is interesting in the OOP time series data and in the previous section on health status from 1974 to 2014, is that time patterns and changes may be reflecting changes in health systems, potentially reflecting impact of changes in how the system is funded or organized.

Figure 4: 20 years trend on OOP spending in select MENA countries

[Insert Figure 4]

Two other countries not charted in Figure 4 but with patterns of health sector OOP worth exploring are Jordan and Lebanon. Lebanon was able to bring down share of OOP on health spending from a high of 60% in 1998 to less than 40% during the last three years; a substantial decline. Jordan’s historic pattern is different, showing first a drastic increase followed by a drastic decline. From 1995 to 20013, OOP spending in Jordan climbed from 23% to over 40%. After that peak, the country has seen a steady decline in the percentage of OOP for health flattening around 22-23% in the last five years. It is important to note that unlike measures of health status, which take long periods of time to move, measures of financial protection, like percentage of OOP spending are more receptive to changes in health systems or to reforms.
Discussion

The fact that health systems are challenging to benchmark due to multiple outcomes and other complicating factors that impact these outcomes could lead to two opposite conclusions. On one hand, we can conclude that benchmarking will be of limited use. On the other hand, and the motivation for this paper and other work on benchmarking health system, is a belief that benchmarking is even more important in this context. If we fail to provide benchmarks, even if through proxy measures, it is impossible to guide future research on lessons from strong or weak systems with the exception of complete failures such as epidemics. That would result in little learning from the vast array of experiences from countries all over the world. Even more concerning if we fail to provide some benchmarks is the inability to then hold sectors and countries accountable for their actions or inactions.

The application of this simple benchmarking exercise to the MENA region’s health sector provided some interesting, and in some cases surprising, conclusions and findings. Different measures of health system performance, representing different dimensions of health system outcomes, do not necessarily move together. In fact, in the MENA case, it appears that the measure of health status (conditioned on the level of socio economic development) appears to be better for countries that performed less well on financial protection (e.g., Morocco and Egypt), and vice versa (e.g., GCC countries). Lebanon and Bahrain are the few countries that achieved better health outcomes relative to income and education, and performed relatively well in provision of financial protection. Some in-depth analysis is warranted for countries that achieve better health outcomes at lower spending than expected, so that their lessons can be learnt by other countries.
On the measure of health outcomes, some countries in the region have consistently outperformed expectations in LEB for a sustained period of time, making those health systems useful targets for research to better understand drivers of success (e.g., Tunisia, Morocco and Egypt). The opposite is also true where some countries have consistently underperformed (e.g., Qatar and UAE). The data on OOP as proxies for financial protection show considerable variations both across countries and within countries over time, making it an excellent resource for detailed assessment of drivers of performance.

Historically, international development agencies have used cross country comparisons and benchmarking to compare countries and to drive the important research agenda described above. A different, and we argue, more important use of benchmarking is to drive the internal reform programs of countries themselves. A successful national reform effort in health usually starts by identifying the goals a country wants to pursue in the health sector. Benchmarking, in this context, is a useful instrument for countries to identify the areas that the reform process needs to address. Knowing, for example, that the country is performing less than expected on health outcomes and out of line with other countries in the region can be used to motivate a reform agenda that address the factors contributing to lower performance. Knowing how a country is performing relative to other countries and when socioeconomic factors are factored in is a strong motivator and driver for health sector reform and strengthening.

Notes
a. IMR = Infant mortality rate; CMR = Child mortality rate; LEB = Life expectancy at birth; 45Q15 = adult mortality or number of people dying between ages 15 and 60.

b. Educational stock is measured by average number of schooling among population aged 15 and above, and from two databases (Barrow-Lee dataset\textsuperscript{9} for about 150 countries and Wittgenstein projections database\textsuperscript{10} for the rest. Barrow-Lee dataset (providing educational attainment measures every five years) has been widely used as a reliable source for educational attainment of countries. The use of Wittgenstein projections database (also for year 2010) for countries not available in Barrow-Lee dataset is justified by the high correlation between the two measurements (r=0.93).
References


Table 1. Background characteristics of MENA countries

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>5,490</td>
<td>75</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Bahrain</td>
<td>20,860</td>
<td>77</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Djibouti</td>
<td></td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egypt, Arab Rep.</td>
<td>3,210</td>
<td>71</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Iran, Islamic Rep.</td>
<td>6,550</td>
<td>75</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Iraq</td>
<td>6,410</td>
<td>69</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Jordan</td>
<td>4,590</td>
<td>74</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Kuwait</td>
<td>49,300</td>
<td>75</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Lebanon</td>
<td>8,120</td>
<td>79</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Libya</td>
<td>7,820</td>
<td>72</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Malta</td>
<td>21,000*</td>
<td>82</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Morocco</td>
<td>3,080</td>
<td>74</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Oman</td>
<td>18,340</td>
<td>77</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Qatar</td>
<td>92,320</td>
<td>79</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>25,500</td>
<td>74</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Syrian Arab Republic</td>
<td></td>
<td>70</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Tunisia</td>
<td>4,160</td>
<td>74</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>44,640</td>
<td>77</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>West Bank and Gaza</td>
<td>3,090</td>
<td>73</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Yemen, Rep.</td>
<td>1,300*</td>
<td>64</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

*Values are for year 2013.
Table 2: Summary of regression results

<table>
<thead>
<tr>
<th></th>
<th>Regression for LEB</th>
<th>Regression for per capita health spending (logged)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observation</td>
<td>166</td>
<td>163</td>
</tr>
<tr>
<td>Goodness of fit</td>
<td>( R^2 = 0.69 )</td>
<td>( R^2 = 0.95 )</td>
</tr>
<tr>
<td>Coefficient for per capita GNI (logged)</td>
<td>3.49*</td>
<td>1.04*</td>
</tr>
<tr>
<td>Standard error for per capita GNI (logged)</td>
<td>0.42</td>
<td>0.03</td>
</tr>
<tr>
<td>Coefficient for average years of schooling</td>
<td>0.76*</td>
<td>0.03</td>
</tr>
<tr>
<td>Standard error for average years of schooling</td>
<td>0.21</td>
<td>0.02</td>
</tr>
</tbody>
</table>

*: \( p < 0.01 \)