Indicators of improved water access in the context of schistosomiasis transmission in rural Eastern Region, Ghana

Alexandra V. Kulinkina a,⁎, Karen C. Kosinski b, Jeanine D. Plummer c, John L. Durant a, Kwabena M. Bosompem d,e, Michael N. Adjei f, Jeffrey K. Griffiths g, David M. Gute a, Elena N. Naumova a,g,h

HIGHLIGHTS

• Safe water access is an important component of integrated schistosomiasis control.
• Defining access indicators at local scales can aid in targeted resource allocation.
• Water source functionality significantly reduces access indicator values.
• Surface water access is negatively correlated with improved water access.
• Correlation is modified by groundwater quality and payment for improved water.

ABSTRACT

Populations with poor access to water, sanitation and hygiene (WASH) infrastructure are disproportionately affected by the neglected tropical diseases (NTDs). As a result, WASH has gained increasing prominence in integrated control and elimination of NTDs, including schistosomiasis. In order to identify underserved populations, relevant measures of access to WASH infrastructure at sub-national or local levels are needed. We conducted a field survey of all public water sources in 74 rural communities in the Eastern Region of Ghana and computed indicators of water access using two methods: one based on the design capacity and another on the spatial distribution of water sources. The spatial method was applied to improved and surface water sources. According to the spatial method, improved water sources in the study area were well-distributed within communities with 95% (CI95%: 91, 98) of the population having access within 500 m when all, and 87% (CI95%: 81, 93) when only functional water sources were considered. According to the design capacity-based method, indicator values were
1. Introduction

1.1. Role of WASH in NTD control

Waterborne and water-related infections constitute a significant burden on population health, particularly in low-income countries. Diarrheal diseases alone account for 1.26 million deaths per year (GBD, 2013), an estimated 58% of which can be attributed to inadequate water, sanitation and hygiene (WASH) (WHO, 2014). Populations with poor access to WASH infrastructure are also disproportionately affected by neglected tropical diseases (NTDs) (Esrey et al., 1991; Grimes et al., 2014; Stocks et al., 2014; Strunz et al., 2014). Hence, WASH has recently gained increased prominence in the context of integrated control and elimination of NTDs, including schistosomiasis (Campbell et al., 2014; Grimes et al., 2015; Steinmann et al., 2006; Spiegel et al., 2010; Utzinger et al., 2003). Spiegel et al. (2010), for example, argue that schistosomiasis control efforts can be enhanced by effective partnerships between the health sector responsible for preventive chemotherapy and the engineering sector responsible for the provision of WASH infrastructure.

1.2. Measures of improved water access

In schistosomiasis endemic areas, availability of sanitation facilities to prevent the contamination of water resources with human waste as well as the availability of improved water sources free from infectious agents are two important components of WASH that affect disease transmission. In the present manuscript we focus on the water rather than the sanitation component; more specifically on identifying populations with poor access to improved water sources in the context of schistosomiasis risk. The WHO and UNICEF’s Joint Monitoring Programme for Water Supply and Sanitation (JMP) defines water from piped water systems, boreholes, protected wells or rainwater harvesting systems as “improved”. Target 7c of the Millennium Development Goals (MDGs) defines “access” as the availability of at least 20 L of water per person per day from an improved source within 1 km (15 min walking time) of the residence (UNDG, 2003). In general, an indicator of access to improved water sources (\(I_{wss} \)) in its simplest form can be expressed as: \(I_{wss} = \frac{P_{wss}}{P_{tot}} \times 100\), where \(P_{wss}\) is the number of people served by improved water sources, and \(P_{tot}\) is the total population. \(P_{wss}\) can be estimated from consumer (i.e. household survey) or provider (i.e. water authority) data.

When using consumer data (such as those offered by the Demographic and Health Survey and country censuses), \(I_{wss}\) is calculated by dividing the number of people who report using an improved water source by the total number of people surveyed. This methodology is used to calculate the JMP indicator, which serves an important purpose of tracking progress towards universal water access over time. However, limitations of the indicator include not fully addressing the accessibility of water services and water quantity and quality (Clasen, 2012). For example, the indicator does not actually account for the walk time or distance to water sources, their functional and seasonal availability, and microbiological or chemical water quality (Kaysen et al., 2013).

Furthermore, indicator values are typically derived at an aggregated level (e.g. JMP provides national estimates stratified by urban and rural locality) while significant differences in water access can exist at regional, district and local levels because the density of both water points and populations may be highly variable across these spatial units (Ntzou et al., 2015). Despite recent methodological advances in mapping sub-national indicator estimates using consumer data (Pullan et al., 2014), substantial differences in indicator values can still exist at the village or sub-village levels due to local heterogeneity in topography, functionality of water sources, and other social, economic, demographic and environmental factors (Bartram et al., 2014). These differences are important to capture for sustainable water infrastructure development as technologies are most often implemented at local scales. Geospatial water source data offered by the water point mapping (WPM) approach (WaterAid, 2016), has been utilized in recent studies (Giné-Garriga et al., 2013; Ntzou et al., 2015) and enables a more detailed evaluation of water access at the local level that is centrally relevant to schistosomiasis control.

When using provider data, in contrast with consumer data, several options of calculating \(I_{wss}\) are available. For example, the indicator can be estimated based on the design capacity of water sources used in Ghana, such as 300 people per standpipe or borehole or 150 people per protected hand–dug well (CWSA, 2014). In this case, \(P_{wss}\) is calculated as the number of improved sources multiplied by the design capacity. Alternatively, if the locations of the water sources are available, such as with the WPM approach, \(P_{wss}\) can be calculated with respect to distance (e.g. 1000 m). The disadvantage of working with provider data is that it is often challenging to define the denominator (\(P_{tot}\)) due to unavailability of appropriate population data. In the case of the design capacity–based approach, \(P_{tot}\) can be obtained from census data disaggregated to the appropriate spatial unit (e.g. community level), which for remote or rural areas may not be readily available. In the case of the distance-based approach, population density data are necessary, which are even more difficult to obtain.

1.3. Ghanaian water context

In rural Ghana, reliance on groundwater is high, with over 95% of improved water supplies intended for domestic purposes being extracted from this source (Awuah et al., 2010). However, high levels of minerals and metals preclude extensive exploitation of groundwater resources in some areas, including the Eastern Region (Rached et al., 1996; Awuah et al., 2010; CWSA, 2016). There is some evidence that boreholes affected by water quality problems such as high concentrations of iron and manganese are either abandoned entirely or used in a limited capacity for purposes other than cooking, drinking, and laundry, which results in wasted investment in the infrastructure (CWSA, 2016; Siabi, 2004).
The majority of donor funds for rural water supplies in sub-Saharan Africa are dedicated to drilling of communal boreholes (Marks et al., 2014); however, in a review of 20 countries it was estimated that at any time, only about 64% of these are functional, ranging from 33% to 90% (Rural Water Supply Network, 2009). In a Ghanaian study, functionality rate of water sources was higher at 79.4% (Fisher et al., 2015). The literature suggests that functionality and sustainability of rural water supplies are enhanced by local revenue generation through water user fees (Rogers et al., 2002; Montgomery et al., 2009). However, payment for water in many rural areas of Ghana is a relatively new concept and willingness to pay remains low due to the belief that water provision is the government’s responsibility and water should be free (Rached et al., 1996; Thorsten, 2007).

In the context of disease risk reduction, the improved water access indicator should account for the consistency with which water sources are used as well as the volume of water collected. Distance, water quality, and price of water are known to influence water use patterns and health benefits offered by improved water sources (Cairncross, 1987; DeGabriele, 2002; Fuest, 2005; Howard and Bartram, 2003; Rogers et al., 2002; Overbo et al., 2016; Pickering and Davis, 2012). Households that self-report using an improved water source for the JMP indicator may still extensively rely on unimproved water sources, especially when the improved source is frequently in disrepair, is of poor perceived water quality for domestic purposes, or is too expensive to use exclusively (Kosinski et al., 2016; Kulinkina et al., 2016). Reliance on untreated surface water, even when temporary and sporadic, places populations at risk for waterborne and water-related diseases. For example, in a rural Eastern Region community, Asamama, surface water use is extensive due to poor groundwater quality and lack of borehole maintenance. As a result, schistosomiasis prevalence in Asamama is >60% despite annual mass drug treatment through the public school system (Kosinski et al., 2016).

1.4. Study goals and objectives

In the present study, we evaluated the availability of public improved water sources at the community level using a distance-based (spatial) and a design capacity–based (non-spatial) approaches. We conducted a field study of all public water sources in 74 rural communities in the Eastern Region of Ghana where surface water is not scarce and schistosomiasis is endemic. Data collection involved water point mapping (WPM); the data were used to meet the following objectives:

1) Estimate access indicators for improved and unimproved water sources using multiple methods and compare the results;
2) Examine the relationship between improved and unimproved water access indicators as modified by payment for water and water quality of improved sources;
3) Compare indicator estimates using two publicly available sources of population density data at ground-truthed field data through a validation study.

Our study intended to promote the development of simple indicators at the local level to provide relevant information for equitable and sustainable water development in rural areas of African countries where water provision is largely decentralized (Giné-Garriga et al., 2013; Thorsten, 2007; WSP, 2011).

2. Methods

2.1. Study area

The Eastern Region is the third most populous of Ghana’s ten administrative regions. Over half of the population (57%) resides in rural areas (GSS, 2012). The primary land cover and land use categories are forests and agricultural land, including hilly areas and low mountain ranges that reach the elevation of approximately 750 m. Rainy season occurs between March and October and is characterized by two peak rainfall periods; the first lasts from May to June and the second from September to October. The dry season lasts from November to February. Pra, Birim, Ayensu and Densu are major perennial rivers that drain the Eastern Region (Fig. 1). Historically, towns located on these rivers exclusively relied on them for their water needs. The Pra and Birim rivers are now polluted by alluvial gold mining activities and are no longer used for domestic purposes due to high turbidity, while the Ayensu and Densu rivers are still used for domestic purposes.

2.2. Sampling strategy

The Eastern Region is divided into 26 administrative districts following the most recent district sub-divisions in 2012. Ten adjacent districts were chosen as the study area because they were similar in their elevation profiles, density of minor rivers and streams, land cover and land use characteristics, and were outside of a 20-km buffer zone of Lake Volta (Fig. 1). Communities situated on the shores of large African lakes such as Lake Volta are historically known to be endemic for schistosomiasis (Onori et al., 1963); however, we were interested in studying exposure constituted by the minor rivers and streams (Gryseels et al., 2006; Kosinski et al., 2016; Salawu and Odaibo, 2016; Ugbomoiko et al., 2010) that are common in tropical parts of West Africa and are not easily detected in satellite images. Although the new (post-2012) district boundaries were used during data collection and all 10 administrative districts were approached, throughout the manuscript, we use older district boundaries in the summary tables to correspond to government data sources that still use older district definitions. We present our results for seven districts: Akyemansa (AKY); Atiwa (ATW); Birim Central (BRC); Birim South (BRS); Kwaebibirem (KBR), which was split in 2012 to be Kwaebibirem and Denkyembour; Suhum-Kraboa Coaltar (SKC), which was split in 2012 to be Suhum and Ayensuano; and West Akim (WAK), which was split in 2012 to be West Akim and Upper West Akim.

To define the universe of towns from which study towns were selected, the most recent settlement–level population data for the Eastern Region from the 2000 Census were obtained from the Ghana Statistical Service (GSS). A data layer of settlement locations was also obtained from the Center for Remote Sensing and Geographic Information Services (CERSGIS). Census data were manually geocoded for all towns within the study districts in the target population range of 500–5000. In Ghana, localities with ≤5000 inhabitants are termed rural (GSS, 2012). Towns with <500 residents were excluded from the study because they were difficult to locate on official maps and difficult to access by public transportation, which was relied upon to conduct the study. A total of 226 towns that met the inclusion criteria were mapped in the 10 districts. Subsequently, the 2000 Census populations were projected to 2014 populations using the 2.1% annual population growth estimated for the Eastern Region by the 2000 and 2010 Censuses (GSS, 2012).

The sampling strategy resulted in a slight over-representation (Table 1) of medium–sized towns in the 1000–2000 and 2000–5000 population categories as classified by the United Nations (United Nations, 2016). Although rural towns are defined as those with populations ≤5000 inhabitants, in our study, we considered towns in the lower end of the 5000–9999 population category as rural due to uncertainties associated with population projection (Table 1).

Ghana Health Service (GHS) Disease Control offices in all 10 districts were consulted for a list of known or suspected schistosomiasis endemic towns, which were also mapped. The definitions of “endemic” varied across districts. Some GHS offices
relied on the results of historical prevalence surveys and limited hospital diagnostic data from the national surveillance system, others relied on lists where praziquantel is currently included in school-based deworming regimens, while the remaining offices simply provided lists of towns where children are known to swim in local freshwater bodies. Of the 226 towns that met our inclusion criteria, 74 (33%) were purposively selected for the study (Tables 1 and 2). Purposive sampling consisted of selecting 'endemic' towns (41/74) and nearby 'non-endemic' towns (33/74) with similar population characteristics. Because no standard definition was applied by district GHS offices to identify endemic towns, endemic status was only used in town selection but not in quantitative analyses.

2.3. Field data collection

Field data were collected between December 2013 and April 2014 by a team of two individuals (including one native Twi speaker) and a local guide in each of the 74 towns between 9 am and 6 pm during all days of the week. In each town, all publicly available and accessible water sources were enumerated and mapped, including public standpipes (SPs) in piped water systems (PWSs), drilled boreholes (BHs), hand dug wells (HDWs), and surface water access points (SWAPs). The local guide was asked to show all SWAPs including those that did not contain water in the dry season. Each source was photographed and its GPS coordinates were obtained using GPS Tracks app (version 2.4.3) for the iPad. In addition, brief surveys were completed about BHs and SWAPs using the QuickTapSurvey app (version 5.4.1) for the iPad. Information was obtained verbally from town residents in Twi and recorded in English.

BH attributes included observed pump type, functionality status, reported payment mechanism, and water quality (WQ) problems. SWAP attributes included observed and reported water uses (fetching, bathing, washing clothes, and swimming). Each SWAP was subsequently classified as ‘perennial’ if it contained water during the dry season field visit. Observed attributes were obtained using GPS Tracks app (version 2.4.3) for the iPad. In addition, brief surveys were completed about BHs and SWAPs using the QuickTapSurvey app (version 5.4.1) for the iPad. Information was obtained verbally from town residents in Twi and recorded in English.

**Table 1**

Distribution of sampled towns among UN population categories according to projected 2014 population data (based on census data from 2000).

<table>
<thead>
<tr>
<th>UN population category</th>
<th>Total mapped towns</th>
<th>Total mapped town population</th>
<th>Total sampled towns</th>
<th>Total sampled town population</th>
<th>% sampled population</th>
<th>% sampled towns</th>
</tr>
</thead>
<tbody>
<tr>
<td>500–999</td>
<td>69</td>
<td>56,130</td>
<td>13</td>
<td>10,907</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>1000–1999</td>
<td>76</td>
<td>108,093</td>
<td>29</td>
<td>45,021</td>
<td>38</td>
<td>42</td>
</tr>
<tr>
<td>2000–4999</td>
<td>64</td>
<td>204,708</td>
<td>27</td>
<td>87,239</td>
<td>42</td>
<td>43</td>
</tr>
<tr>
<td>5000–9999</td>
<td>17</td>
<td>96,336</td>
<td>5</td>
<td>28,610</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>226</td>
<td>465,267</td>
<td>74</td>
<td>171,777</td>
<td>33</td>
<td>37</td>
</tr>
</tbody>
</table>

Fig. 1. Map of the study area.
BH survey respondents were asked to state whether and how users pay for water from the BH and to comment on any WQ problems the users experience in an open-ended format. The advantage of this method is that responses are not constrained to a concrete list. The disadvantage is that if not specifically prompted, the user may have forgotten to list all problems. Open-ended format was preferred because we had no prior knowledge of WQ in the study area and we wanted to encourage a wide range of responses. SWAP survey respondents were asked to respond “yes” or “no” to whether or not the source is used by the community for each of the following: fetching water, bathing, washing clothes and swimming. The study was excluded from ethical review because it involved collection of data about individuals. The study was excluded from ethical review because it involved collection of data about individuals.

The spatial method was applied to all three indicators. The indicator value (% population with access within a predefined distance) was approximated as the % of the town area (proxy for % population) within the buffer and calculated as follows:

\[ I_{\text{buffer dist}} = \frac{A_{\text{buffer}}}{A_{\text{total}}} \times 100 \]

where \( A_{\text{buffer}} \) is the area within the predefined buffer of the water source and \( A_{\text{total}} \) is the total area of the town represented by the town polygon abstracted from Open Street Map using QGIS software (version 2.12.3); uniform population density was assumed throughout the polygons. A total of 18 indicators were derived using the spatial method: \( I_{\text{iws}(100-500, 1000)} \), \( I_{\text{iws}(100-500, 1000)} \), and \( I_{\text{psws}(100-500, 1000)} \).

To assess the impact of functionality on the indicator values, the differences between \( I_{\text{iws}} \) and \( I_{\text{psws}} \) values for all relevant permutations of the analysis were tested for significance using a paired t-test on the full dataset \( n = 74 \) as well as within each district \( n = 7 \) and population category \( n = 4 \). District and population category averages were also compared to the overall average using a Welch two sample t-test, the results of which are valid with unequal variances and unevenly distributed sample sizes.

To achieve the second objective of examining the relationship between improved and surface water access, Spearman’s rank correlations between \( I_{\text{iws}} \) or \( I_{\text{iws}} \) and \( I_{\text{psws}} \) were explored on the full dataset \( n = 74 \) as well as on subsets of the data defined by payment mechanism and reported WQ. To define the subsets, towns were divided into two groups based on whether they did \( n = 38 \) or did not \( n = 36 \) have a regular proactive payment mechanism for water from improved sources (defined as per use or per unit of time payment); and into three groups by whether they had no reported WQ problems \( n = 25 \), one problem \( n = 31 \) and two or more problems \( n = 18 \). Further information about payment and WQ are provided in Section 3.1.2. After examination of the correlation matrix, \( I_{\text{psws}(300)} \) was modeled as a function of \( I_{\text{iws}} \) controlling for population, payment, and WQ problems using generalized linear regression. No interactions between the variables were statistically significant.

To achieve the third objective of validating the spatial indicator method, \( I_{\text{iws}} \) values derived using Open Street Map town polygons were compared to those derived with population

### Table 2
Population statistics for the study area, universe of potential study towns, and study towns according to projected 2014 population data (based on census data from 2000).

<table>
<thead>
<tr>
<th>Study area (7 districts)</th>
<th>AKY</th>
<th>ATW</th>
<th>BRC</th>
<th>BRS</th>
<th>KBR</th>
<th>SKC</th>
<th>WAK</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population</td>
<td>105,815</td>
<td>120,211</td>
<td>157,427</td>
<td>130,149</td>
<td>209,254</td>
<td>182,075</td>
<td>212,282</td>
<td>1,117,213</td>
</tr>
<tr>
<td>Rural population</td>
<td>68,908</td>
<td>80,116</td>
<td>50,884</td>
<td>68,302</td>
<td>107,038</td>
<td>134,183</td>
<td>130,499</td>
<td>639,929</td>
</tr>
<tr>
<td>% rural population</td>
<td>65.1</td>
<td>66.6</td>
<td>32.3</td>
<td>52.5</td>
<td>51.2</td>
<td>73.7</td>
<td>61.5</td>
<td>57.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mapped towns within study area (226 towns)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Towns (#)</strong></td>
</tr>
<tr>
<td>Total population</td>
</tr>
<tr>
<td>Median population</td>
</tr>
<tr>
<td>% of study area rural population</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study towns (74 towns)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Towns (#)</strong></td>
</tr>
<tr>
<td>Total population</td>
</tr>
<tr>
<td>Median population</td>
</tr>
<tr>
<td>% of mapped towns</td>
</tr>
<tr>
<td>% of mapped town population</td>
</tr>
<tr>
<td>% of study area rural population</td>
</tr>
</tbody>
</table>

where \( P_{\text{tot}} \) is the total population of each town obtained from the census. A total of two indicators were derived using the non–spatial method: \( I_{\text{iws}} \) and \( I_{\text{psws}} \).

The spatial method used the GPS coordinates of water sources with buffer distances of 100, 200, 300, 400, 500, and 1000 m drawn in ArcGIS software (version 10.2.2) based on evidence from existing literature (Howard and Bartram, 2003; CWSA, 2014; UNDG, 2003). The indicator value (% population with access within a predefined distance) was approximated as the % of the town area (proxy for % population) within the buffer and calculated as follows:

\[ I_{\text{buffer dist}} = \frac{A_{\text{buffer}}}{A_{\text{total}}} \times 100 \]

where \( A_{\text{buffer}} \) is the area within the predefined buffer of the water source and \( A_{\text{total}} \) is the total area of the town represented by the town polygon abstracted from Open Street Map using QGIS software (version 2.12.3); uniform population density was assumed throughout the polygons. A total of 18 indicators were derived using the spatial method: \( I_{\text{iws}(100-500, 1000)} \), \( I_{\text{iws}(100-500, 1000)} \), and \( I_{\text{psws}(100-500, 1000)} \).

To assess the impact of functionality on the indicator values, the differences between \( I_{\text{iws}} \) and \( I_{\text{psws}} \) values for all relevant permutations of the analysis were tested for significance using a paired t-test on the full dataset \( n = 74 \) as well as within each district \( n = 7 \) and population category \( n = 4 \). District and population category averages were also compared to the overall average using a Welch two sample t-test, the results of which are valid with unequal variances and unevenly distributed sample sizes.

To achieve the second objective of examining the relationship between improved and surface water access, Spearman’s rank correlations between \( I_{\text{iws}} \) or \( I_{\text{iws}} \) and \( I_{\text{psws}} \) were explored on the full dataset \( n = 74 \) as well as on subsets of the data defined by payment mechanism and reported WQ. To define the subsets, towns were divided into two groups based on whether they did \( n = 38 \) or did not \( n = 36 \) have a regular proactive payment mechanism for water from improved sources (defined as per use or per unit of time payment); and into three groups by whether they had no reported WQ problems \( n = 25 \), one problem \( n = 31 \) and two or more problems \( n = 18 \). Further information about payment and WQ are provided in Section 3.1.2. After examination of the correlation matrix, \( I_{\text{psws}(300)} \) was modeled as a function of \( I_{\text{iws}} \) controlling for population, payment, and WQ problems using generalized linear regression. No interactions between the variables were statistically significant.

To achieve the third objective of validating the spatial indicator method, \( I_{\text{iws}} \) values derived using Open Street Map town polygons were compared to those derived with population.
density raster data (100 m² grid) abstracted from WorldPop (The WorldPop Project, 2016; Linard et al., 2012) and to those derived using GPS coordinates of households collected in June 2016 as part of a separate study (unpublished). This validation study was conducted for 8 of the study towns representing a range of population sizes (two towns per population category). A potential advantage of using WorldPop data as compared to Open Street Map data is that they represent population density (i.e., uniform population density assumption is not necessary as with Open Street Map polygons); however, the spatial resolution of these data is very coarse. Using WorldPop data, $I_{\text{HDWs}}$ was calculated as the sum of population grids intersecting each buffer distance (using zonal statistics; Spatial Analyst extension of ArcGIS) and expressed as a percentage of the total town population obtained from the census. Mapped households represent the most accurate population density data that can be obtained at the community level short of enumerating individuals within households. Using household data, $I_{\text{HDWs}}$ was calculated as the sum of points falling completely within each buffer distance as a percentage of the total number of households in the town. $I_{\text{HDWs}}$ values for the three population data sources were plotted against buffer distance and visually compared.

3. Results

3.1. Characteristics of public water sources

3.1.1. Piped water systems

A total of 88 public SPs were mapped in 7 of 74 towns (9.5%) that had mechanized PWSs supplying groundwater through a network of metered public, private and institutional connections. Of these, 75 SPs (85.2%) were functional and in use at the time of the study. For all PWSs, a nearby resident was hired by the town to sell water from the public standpipes for several hours per day in the mornings and evenings, charging water consumers on a per-use volumetric basis. More information about PWSs in the study area is found in a prior publication (Kulinkina et al., 2016).

3.1.2. Boreholes

A total of 238 BHs were mapped in 67 of 74 towns (90.5%); the other 7 towns had no BHs. Pump types included India Mark II (144/238, 60.5%), Afrifield (52/238, 21.8%), Nira (27/238, 11.3%), and mechanized with multiple spouts (6/238, 2.5%). For 9 BHs (3.8%), pump types were unknown because they were removed due to prolonged disrepair. For mechanized BHs, each spout was considered as an individual water source for a total of 255 spouts. Of the 238 BH locations, 176 had at least one functional spout during the field visit for a total of 190 of 255 (74.5%) functional spouts.

Multiple financing mechanisms existed for BHs. Of 238 BHs, 130 (54.6%) had proactive payment associated with their use, 87 (36.6%) had retroactive payment, 19 (8.0%) BHs were free of charge and for 2 BHs (0.8%) payment type was not available. Proactive payment mechanisms included per-use payment (39.5%) or per unit of time payment (15.1%), most often monthly. Retroactive mechanisms included withdrawing money from a communal fund (26.9%) or collecting money from nearby users (9.7%) once a BH was already broken.

Open-ended questions about BH WQ concerns showed five recurring problems: salty taste, unfavorable scent, oil formation on the water surface, presence of particles, and the water's tendency to stain sandy soils (e.g., plantain, cassava, rice) black/purple color during cooking. Upon derivation of this complete list from the responses, a binary presence or absence variable for these five problems was assigned to each BH for which reported WQ data were available (203/238). Presence of particles was the most prevalent WQ problem (33/203, 16.3% of BHs), followed by unfavorable scent (24/203, 11.8%), oil formation (19/203, 9.4%), salty taste (18/203, 8.9%) and food staining (15/203, 7.4%). Overall, 27% (55/203) of the BHs had at least one reported WQ problem.

3.1.3. Hand-dug wells

Overall, 112 public HDWs were mapped in 35 of 74 the study towns (47.3%). Of these, 39 (34.8%) were covered and fitted with a pump. The others were open wells from which water was withdrawn by lowering a rope with a bucket attached down the well. Uncovered HDWs included those that were covered at one point, but at the time of the study the pump had been removed. Most of the HDWs were free of charge; few collected a smaller fee than what was charged for BH water. Payment mechanism and WQ data were not collected systematically from HDWs.

3.1.4. Surface water

A total of 205 SWAPs were mapped in 71 of 74 towns (95.9%); 3 towns lacked access to surface water. Of these, 186 (90.7%) were perennial. Fetching water for domestic purposes was reported in 194 of 205 (94.6%) locations, swimming in 120 (58.5%), bathing in 73 (37.6%) and washing clothes in 41 (20.0%). Surface water use was observed in 83 of 205 (42.0%) locations.

3.2. Indicator estimates

3.2.1. Non-spatial design capacity-based method

According to the design capacity-based method, the overall average $I_{\text{WQ}}$ value was 63% (CI95%: 57, 69) with the corresponding lower $I_{\text{WQ}}$ value of 49% (CI95%: 43, 55). Examined by district, only in ATW both $I_{\text{WQ}}$ and $I_{\text{HDWs}}$ values were significantly higher ($p < 0.05$) than the overall averages: 82% (CI95%: 71, 93) and 66% (CI95%: 52, 81), respectively (Table S1 in Supporting Information). The average difference between $I_{\text{WQ}}$ and $I_{\text{HDWs}}$ values of 14% was statistically significant ($p < 0.001$), indicating that functionality significantly affects improved water access. Furthermore, the effect of functionality was significant in nearly all of the districts with the exception of AKY and BRC, for which differences between $I_{\text{WQ}}$ and $I_{\text{HDWs}}$ values were not significant most likely due to low sample size. Functionality also significantly affected water access in the two middle population categories (1000–1999 and 2000–4999) (Table S1 in Supporting Information).

3.2.2. Spatial distance-based method

According to the distance-based method, improved water access within 100 m was relatively low, with an average $I_{\text{WQ}(100)}$ value of 22% (CI95%: 19.0, 24) and $I_{\text{HDWs}(100)}$ value of 18% (CI95%: 15, 20). However, the percentages increased fairly rapidly through the distance bands with $I_{\text{WQ}(200)}$ value of 79% (CI95%: 75, 84) and $I_{\text{WQ}(500)}$ value of 95% (CI95%: 91, 98). The corresponding $I_{\text{WQ}(1000)}$ and $I_{\text{HDWs}(1000)}$ values were 70% (CI95%: 63, 76) and 87% (CI95%: 81, 93), respectively. Access within 1000 m was nearly universal with $I_{\text{WQ}(1000)}$ value of 99% (CI95%: 96, 100) and $I_{\text{HDWs}(1000)}$ value of 93% (CI95%: 87, 99) (Table S1 in Supporting Information). When examined by district, AKY and BRC had above-average $I_{\text{WQ}(1000)}$ values; ATW had above-average $I_{\text{WQ}(500)}$ and $I_{\text{HDWs}(1000)}$ values; KBB had above-average indicator values for most buffer distances. BRS had below-average $I_{\text{WQ}(1000)}$ and $I_{\text{HDWs}(1000)}$ values. When examined by population category, above-average $I_{\text{WQ}(1000)}$ values were observed in the two highest population categories (2000–4999 and 5000–9999) (Table S1 in Supporting Information).

The average differences between $I_{\text{WQ}}$ and $I_{\text{HDWs}}$ values determined by the spatial method were lower (4–9%) as compared to those determined by the non-spatial method (14%), but were still significant for all buffer distances ($p < 0.05$) in the full
dataset. When examined by district, primarily ATW and WAK exhibited significant differences \((p < 0.05)\) between the \(I_{\text{IWS}}\) and \(I_{\text{PSWS}}\) values and only for the shorter buffer distances (100–400 m). When examined by population category, only the middle two population categories (1999–2000 and 2000–1999) exhibited significant differences between \(I_{\text{IWS}}\) and \(I_{\text{PSWS}}\) values for nearly all buffer distances (Table S1 in Supporting Information).

Surface water access was significantly lower than improved water access as assessed by the differences between \(I_{\text{IWS}}\) or \(I_{\text{PSWS}}\) and \(I_{\text{PSWS}}\) \((p < 0.001)\) for distances 100–500 m (results not shown). On average, 7% \((C_{\text{IWS}}: 5, 8)\) of the population had access to a perennial surface water source within 100 m, 37% \((C_{\text{IWS}}: 30, 43)\) within 300 m, 62% \((C_{\text{IWS}}: 54, 71)\) within 500 m and 85% \((C_{\text{IWS}}: 77, 92)\) within 1000 m. Significantly higher than average \(I_{\text{PSWS}}\) values for all buffer distances were observed in WAK. Below-average \(I_{\text{PSWS(100)}}\) values were observed in ATW and KBR, \(I_{\text{PSWS(100–200)}}\) in BRC, and \(I_{\text{PSWS(100–300)}}\) in AKY and BRS. Below-average \(I_{\text{PSWS(100–300)}}\) values were also observed in the third population category (2000–1999) (Table S1 in Supporting Information).

### 3.2.3. Correlation between improved and surface water indicators

Spearman’s rank correlations between \(I_{\text{IWS}}\) or \(I_{\text{PSWS}}\) and \(I_{\text{PSWS}}\) were examined on the full dataset and on subsets based on payment mechanism and reported WQ problems (Table S2 in Supporting Information). In the full dataset, relatively low correlation coefficients \((-0.24 \text{ to } -0.28; p < 0.05)\) were observed between \(I_{\text{IWS}}\) and \(I_{\text{PSWS(200–500)}}\). In the subset of towns with no regular payment for water, correlations between \(I_{\text{IWS}}\) and \(I_{\text{PSWS}}\) were stronger \((-0.39 \text{ to } -0.43; p < 0.05)\) and evident for all buffer distances. Correlations also emerged in this subset between \(I_{\text{IWS}}\) and \(I_{\text{PSWS}}\). No significant negative correlations were observed for towns with proactive payment mechanisms for water. Similarly, correlations were examined by WQ. Only in the subset of towns with one problem \((n = 31)\) there were significant negative correlations \((-0.36 \text{ to } -0.38; p < 0.05)\) between \(I_{\text{IWS}}\) and \(I_{\text{PSWS(200–500)}}\), but not for towns with no problems or for towns with two or more problems.

The association between \(I_{\text{IWS}}\) as estimated by the non-spatial method and \(I_{\text{PSWS(300)}}\) was more formally investigated using a generalized linear regression model, controlling for population, payment and WQ (Table 3). The results of the model confirmed the negative relationship between the two indicators, demonstrating that surface water access decreases by 0.3% for each 1% increase in functional improved water access \((p < 0.05)\), controlling for the other parameters. Population size also had a statistically significant negative effect on the \(I_{\text{PSWS(300)}}\) indicator value suggesting that larger towns has less access to surface water \((p < 0.01)\). Having one WQ problem, however, had a very large effect, increasing surface water access by 17% as compared to having no problems \((p < 0.05)\). Payment mechanism was not significant after controlling for the other variables. The model explained 24% of the variability in \(I_{\text{PSWS(300)}}\) indicator values. It should be noted that although surface water access (as a proxy of surface water use) is chosen as the outcome variable in the model, the relationship is bidirectional and does not imply causality.

### 3.3. Population density data sources validation study

The validation study (Fig. 2) showed that in seven out of eight cases, the WorldPop dataset resulted in significantly underestimated \(I_{\text{IWS}}\) value, whereas in all 8 cases, indicator values determined using Open Street Map and GPS coordinates of the households were very close to each other. It should be noted that the analysis is limited by the census population serving as the denominator in the calculation of the indicator using the WorldPop dataset. This limitation cannot be addressed because the true denominator population in this case is not possible to determine without imposing artificial town boundaries onto the raster dataset.

### 4. Discussion

#### 4.1. Major findings

We utilized community level provider water source data (similar to Giné-Garriga et al., 2013 and Ntozini et al., 2015) to compute improved water access indicators using a design capacity–based and a distance–based method for six buffer distances ranging from 100 to 1000 m. Although basic access according to the MDGs is constituted by a water source located with 1000 m of the home (UNDG, 2003), research suggests that optimum water use for good hygiene behavior and better health is achieved when an on-plot water source is available with water use dropping substantially when distance to water source exceeds 100 m (Cairncross, 1987; Curtis et al., 1995; Howard and Bartram, 2003; Larson et al. 2006; Overbo et al., 2016; Pickering and Davis, 2012; Thompson et al., 2001). In our study area, water access within 100 m was rather low \((18–22\%)\) and contributed mostly by towns with PWSs with a higher average density of public SPs than that of BHs. However, the percentages increased rapidly through the distance bands with 87–95% of the study population having access within 500 m, and nearly universal access within 1000 m, indicating a uniform spatial distribution and relatively short walking distances to improved sources within the study towns. Design capacity-based indicator values were lower at 63% when all, and 49% when only functional water sources were considered, indicating that the overall low number of functional sources may create long lines and waiting times, potentially negating the benefit of short walking distances.

Furthermore, improved water access was substantially affected by functionality. By district, functionality had a significant effect in all districts according to the non–spatial method and in ATW and WAK districts according to the spatial method. Concurrently, WAK district had significantly higher than average surface water access in all distance buffers, partially due to the presence of Ayensu River that had not yet been polluted by gold mining at the time of the study. By population, only the two middle categories \((1000–1999\) and \(2000–4999\) people) had significantly lower \(I_{\text{IWS}}\) and \(I_{\text{PSWS}}\) than values according to both spatial and non–spatial methods. The findings are consistent with personal observations. Very small towns \(<1000\) people) typically have few water sources; hence they are motivated to keep them functional. Large towns \(\geq5000\) people) typically have a higher water source density and are less affected by few non–functional sources, particularly when using the spatial method. Hence the observation that indicator values are most influenced by functionality in the two middle categories.

Overall, surface water use is extensive in the study area; 60% of the study population had access to a perennial surface water
source within 500 m and 95% of the mapped sources were used for fetching water for domestic purposes. There was a stable inverse relationship between Iwss and Ipws indicating that higher access to improved sources could potentially reduce reliance on surface water. This relationship, however, is modified by payment and WQ of the improved water sources. The relationship between the indicators was stronger in towns with no regular proactive payment for groundwater sources and not significant in towns with regular payment. Similarly, when stratified by WQ, the inverse relationship was only significant in the subset of towns with one reported problem but not in towns with no problems or with two or more problems. This suggests that while in general, higher access to improved sources may be associated with lower surface water use, this may not necessarily occur in the presence of payment for water or when groundwater quality is perceived to be poor for domestic purposes as compared to surface water. WQ was a more important factor than price in the regression analysis.

4.2. Indicator methodology

We explored two methods of computing indicators and multiple sources of population density data required for the spatial method. The two methods produced somewhat different results (e.g. ATW district had higher indicator values according to the non-spatial method and KBR district had higher values according to the spatial method). It should be noted that the two methods measure different aspects of water access and exhibit tradeoffs in their computation in that neither uses all available information. The non-spatial method uses the total number of spouts (i.e. mechanized BHs can have multiple spouts) but doesn’t use the distance and hence is an indicator of crowding around water sources and long waiting times. The spatial method uses the location of the water source in relation to population density but does not use the total number of spouts and hence is an indicator of walking distance. Both waiting time and walking distance are important determinants of access (Asaba et al., 2013; Gross et al., 2013).
A major drawback of the spatial method is the need for population density data, which are difficult to obtain for rural areas. The validation study showed that while WorldPop raster data are very valuable for analyses conducted at the national level, the Open Street Map town polygons, even under the uniform population density assumption, provide much more similar indicator values to those obtained using ground-truthed GPS locations of households. This was true for small towns of <1000 people as well as large towns of ≥5000 people. It is important to note that the true denominator could not be determined for the analysis that utilized WorldPop data, which is a limitation of this assessment. Because the town polygons are easy to extract from a public online data source and to visually verify and adjust using high resolution Google Earth imagery, they show promise for community level water access indicator development in other locations.

The approaches described here use geocoded provider data to more accurately quantify water access at the local level. Our study encourages the district water and sanitation authorities in Ghana to undertake the effort of geocoding improved and unimproved water sources and tracking their functionality and uses as a step towards further developing water infrastructure aimed at achieving consistent use and disease risk reduction. Although the design capacity–based indicator estimates are still most common at the district level, which rely only on tracking the total number of improved water sources per community (9 out of 10 study districts), one district had GPS coordinates of water sources available. Recent publications suggest that spatial data with more attributes are not far out of reach as Ghana is on a positive trajectory towards establishing an inventory of rural and small-town water supplies with the use of mobile phones (JRC, 2014). When collecting field data about water points, additional attributes such as functionality, seasonality, payment mechanism, and WQ should be incorporated. Our study findings suggest that to reduce reliance on surface water where it is readily available, a sufficient number of improved water sources in close proximity is important, but authorities must also ensure acceptable WQ and payment structure.

4.3. Study limitations and future directions

In terms of data collection, our study has three potential limitations. The first limitation is the use of a single guide per community to map all available water sources. While every effort was made to find a knowledgeable person, it is possible that some guides were more familiar with the community layout than others. A limited validation of the approach was possible from the results of a subsequent survey of BHs in 2015. Using a complete list of BHs discovered during both surveys and excluding newly constructed BHs, mapping accuracy was determined at ~95% for both surveys (unpublished data).

The second limitation is that the representativeness of some of the observed or reported variables may be affected by the day of the week or time of the day when data were collected. For example, people typically fetch water more during mornings and evenings, meaning that towns that were visited during these prime water fetching hours contained information from more users and are expected to be more representative. A third related limitation is in the way that payment mechanism and WQ problems were documented. Because responses were obtained from a small number of respondents in each town, our data is likely not representative of all users. Despite this limitation, we believe that the most prevalent and severe WQ problems were represented by this method, and the method can be widely and inexpensively applied to areas where no water quality information is available. Furthermore, a limited analysis of reported WQ problems and measured WQ parameters for salty taste of groundwater showed that in towns where users ubiquitously noted a salty taste, total dissolved solids, chloride and sulfate concentrations were elevated in the water samples (Kulinkina et al., 2016).

In terms of data analysis and indicator development, the following limitation should be noted. Although in the validation study, the uniform population density assumption within the Open Street Map polygon was reasonable for the range of populations tested and the results compared well with those derived using the mapped households, this assumption is not necessarily valid for all town structures. For example, in our study, very remote communities with small total populations and households that are likely dispersed over a large area were excluded. Therefore, our study findings do not necessarily apply to these communities. Recent advances in rural remote population density mapping that utilize roof reflectance data obtained from remote sensing imagery (Guo et al., 2016; Varshney et al., 2015) demonstrate a promising alternative to the Open Street Map polygon method.

5. Conclusions

Our first main finding is that the functional water coverage was significantly lower than what might be expected from all mapped water sources. Our second main finding was that while water access within 100 m of residence was low, indicator values increased rapidly through the distance bands with nearly universal access within 500 m and 1000 m indicating a uniform distribution of water sources within communities. Thirdly, we found that surface water use was extensive in the study area in the presence of proximal improved water sources. An inverse relationship was observed between improved water access and surface water access, and was modified by the presence of payment and perception of poor groundwater quality of improved sources. Although access is often intended as a proxy for use, in our study, this assumption is only reasonable for surface water sources, but not for improved water sources, particularly due to interactions with payment and water quality. We suggest that geocoded provider source data that incorporate information on functionality, payment and water quality would allow for calculation of more representative water access indicators relevant at sub-national and local scales.

Mapping and indicator computation should be undertaken at the district administrative level to support decision making in the context of decentralized water provision schemes. Furthermore, we suggest that local knowledge of infrastructure functionality rates, prevailing payment mechanisms and perceived water quality concerns would likely lead to infrastructure improvements that will be maintained and utilized, resulting in a more long-lasting impact on population health and on the rates of NTDs than that achieved through mass drug treatment alone.

Author contributions

AVK, KCK, ENN, JKG and DMG designed the study. AVK and MNA collected the data. AVK and ENN analyzed the data. AVK drafted the manuscript. All authors have read and approved the manuscript.

Conflicts of interest

The authors declare no competing interests.
Acknowledgments

This study was funded by the National Institutes of Health (R34 AI097083-01A1) and Tufts Institute for Innovation grants. We wish to acknowledge members of Ghana Health Service, Ghana Statistical Service and Community Water and Sanitation Agency (particularly Gilbert A. Ayamgah and Theophilus Mensah) for the provision of various datasets, and members of the traditional leadership of the study communities for allowing us to conduct mapping and survey activities. We are thankful to Miguel Stadecker for providing valuable input to the manuscript and to Michelle Sedipo, Olivia Schultes, Isaac Kwaku Cudjoe, Kezia Osei Bonu and Emmanuel Asare Agyapong for assisting with the mapping of households for the validation study. We also appreciate the helpful comments and suggestions of the Science of the Total Environment reviewers and editor.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.scitotenv.2016.11.140.

References
