Research Paper

Cancer risk assessment and modeling of groundwater contamination near industrial estate, Lahore, Pakistan

Muhammad Irfan Jalees, Muhammad Umar Farooq and Asma Tufail Shah

ABSTRACT

More than 60% of the population of Pakistan has no access to safe drinking water. Industrial zones near populated areas make conditions more severe due to continuous contamination. The aim of this study was to use statistical tools for correlation and source identification and health risk assessment of contamination due to Sundar Industrial Estate (SIE), Lahore, Pakistan. Drinking and wastewater samples were collected from SIE and analyzed for physical, chemical, microbial, and heavy metals analysis. Results showed that heavy metals and microbial contamination were beyond the National Drinking Water Quality Standards of Pakistan while high values of chemical oxygen demand (COD) and biochemical oxygen demand (BOD) wastewater were responsible for contamination of drinking water through seepage. There was a medium to strong correlation among parameters of all samples as indicated by Pearson correlation and analysis of variance. Principal component analysis and cluster analysis indicated sources of contamination, i.e., refuse leachate and untreated effluent discharges as main source of pollutants for drinking water. Health risk assessment showed a high intake of heavy metals through drinking water. Hazard quotient and hazard index indicated high probability of non-carcinogenic risk while cancer risk assessment suggested that out of every 100 of the population 93 people may suffer carcinogenic effects.

Key words | correlation, heavy metals, risk assessment, source, statistical modeling

Muhammad Irfan Jalees (corresponding author)
Muhammad Umar Farooq
Institute of Environmental Engineering and Research,
University of Engineering and Technology,
Lahore 54890,
Pakistan
E-mail: *irfan611@gmail.com*

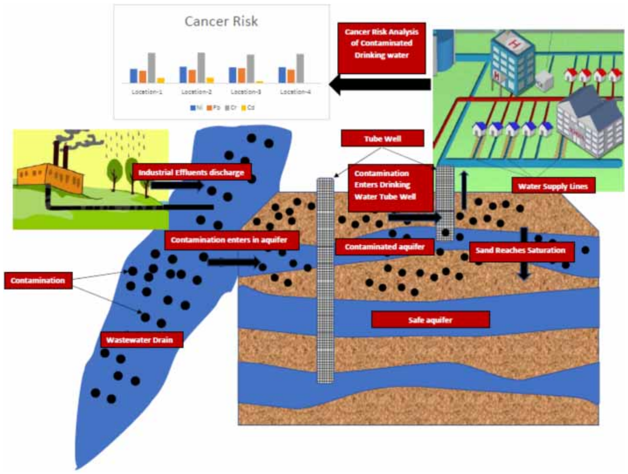
Asma Tufail Shah
Interdisciplinary Research Centre in Biomedical Materials,
COMSATS Institute of Information Technology,
Lahore,
Pakistan

HIGHLIGHTS

- Samples showed high bacterial and heavy metals contamination due to untreated effluent.
- Statistical modeling showed strong correlation and seepage as the main source of pollution.
- The hazard quotient and hazard index (for non-carcinogenic risk) values are >1.
- Risk assessment showed 93 people out of 100 of population may suffer from cancer.

doi: 10.2166/washdev.2021.217

GRAPHICAL ABSTRACT



INTRODUCTION

Water is an essential element for life. A small proportion (0.01%) of fresh water is available for human use (Azizullah *et al.* 2011). Unfortunately, even this small proportion is continuously contaminated by various anthropogenic sources including urbanization/industrialization (Rehman *et al.* 2008; Valipour 2016). Like other developing countries, Pakistan is also facing a problem of contamination in drinking water due to anthropogenic sources. A recent study showed drinking water in Pakistan (4,218 samples) was contaminated by bacteria (69%), chemicals (19%), and heavy metals (24%) (Saiqa *et al.* 2016). Drinking water can be contaminated by natural or anthropogenic sources. Various studies have indicated that natural sources like the flood (in Pakistan) of 2010 and earthquakes of 2005 and 2008 contaminated the groundwater with microbes (pathogens) (Baig *et al.* 2012a, 2012b; Khan *et al.* 2014; Saeed & Attaullah 2014). The main sources of anthropogenic activities are industries and untreated wastewater discharge in rivers and canals (Azizullah *et al.* 2011). Such anthropogenic activities can be divided into point sources and non-point (diffused) sources (Rehman *et al.* 2008; Azizullah *et al.* 2011; Raza *et al.* 2017). Industrial and domestic effluents are categorized as point sources while runoff of agriculture and hard surfaces are non-point sources (Raza *et al.* 2017; Valipour 2017).

In Pakistan, out of 6,634 registered industries, 1,228 were considered highly polluted point sources for the environment (Raza *et al.* 2017). These industries are named as Small Industries-II, Gujranwala City; Industrial Estate Peshawar, Peshawar City; Industrial Estate Hattar; Quid-E-Azam Industrial Estate, Lahore City; Sundar Industrial Estate, Lahore City, etc. Like every developing city, Lahore is also facing problems of safe drinking water due to these point sources. People are drinking this polluted water and facing severe health issues leading to cancer. This study aimed to check the quality and drinking water and pollutant correlation along with health risk assessment due to drinking contaminated water. Sundar Industrial Estate of Lahore, Pakistan was selected, and sampling of drinking water and wastewater was performed. Statistical tools like descriptive statistics, ANOVA, PCA, CA, etc. were used to calculate correlation and identify the sources of pollutants present in the study area. Health risk assessment as hazard quotient (HQ) and cancer risk assessment (CR) was also determined.

MATERIALS AND METHODS

Sundar Industrial Estate (SIE) is in Lahore City, with an area of 1,800 acres (Figure 1). SIE was established in 2007, and currently, more than 550 industrial units are working.

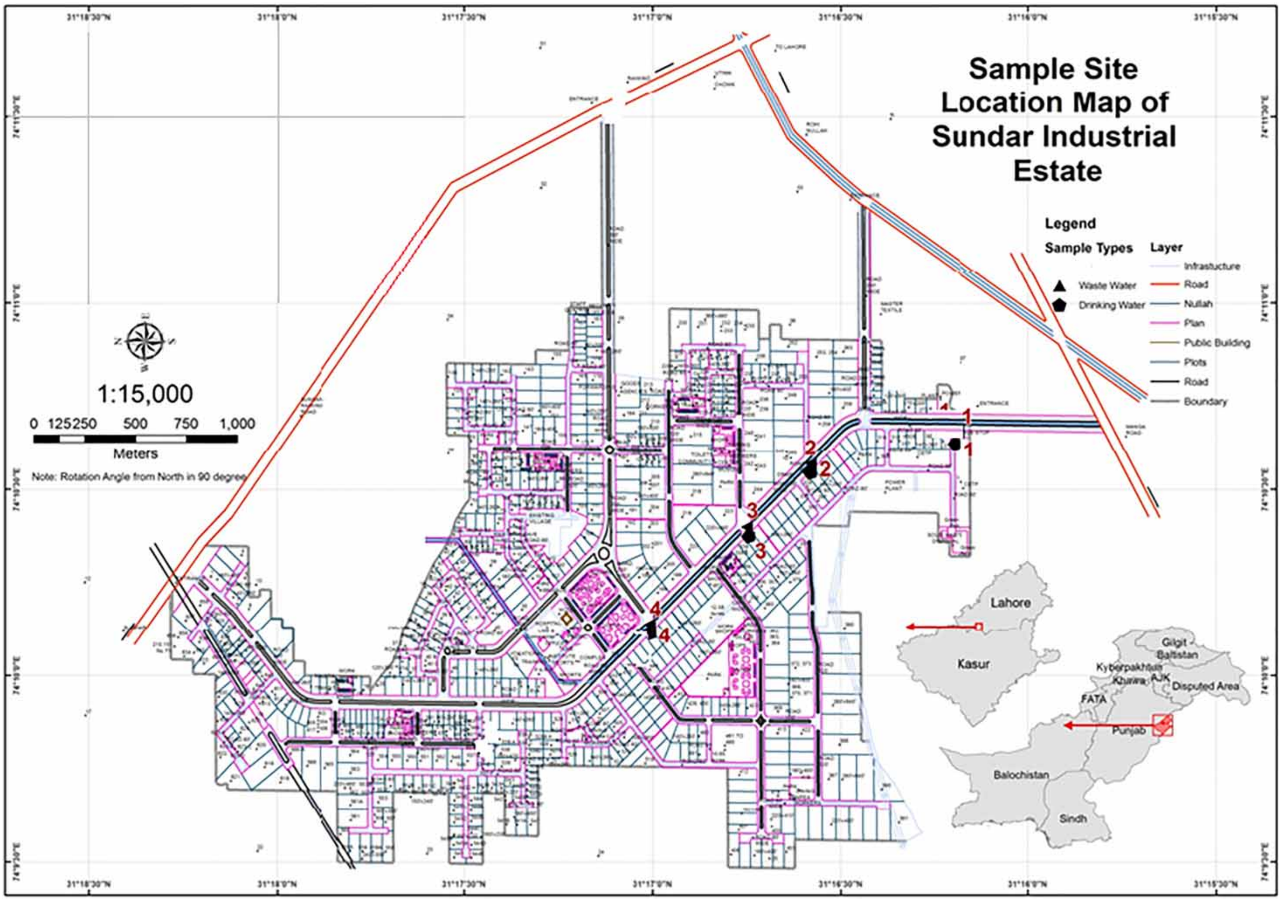


Figure 1 | Location map of Sunder Industrial Estate. Locations of drinking and wastewater are labeled as 1, 2, 3, and 4.

The main industries are given in Table 1. The Environmental Impact Assessment (EIA) report of SIE (PIEDMC 2006) indicated a total of 140,600 m^3/day wastewater will be produced which needs installation of a wastewater treatment plant with a capacity of 150,000 m^3/day (PIEDMC 2006). Unfortunately, this treatment facility was never built. This abundant untreated wastewater contaminates the drinking water of SIE and has resulted in poor health conditions for the population living near SIE.

Sampling and analysis

The wastewater samples were collected from the main drain of SIE (Figure 1) and drinking water samples were collected from taps (home) near the wastewater sample locations. Four locations were selected for wastewater and drinking water sampling. Five samples from each wastewater and drinking water location were collected at an interval of 2 days. Water sampler (WS700) was used for the collection of composite sampling (APHA 2005). The sampler was adjusted to take 1 L of wastewater sample after every 30 min. A total of 20 L wastewater samples was collected from each location and was repeated after every 2 days' interval. For drinking water sampling, the tap water was collected. The water taps were left running for 10–15 min and, after that, a 1 L sample was collected after every 30 min. The same was repeated at every drinking water location with a 2-day interval of time. Samples were collected in sterilized dark glass bottles to avoid any degradation and bacterial contamination. Afterwards collection samples were transferred to the laboratory for analysis (APHA 2005). Samples were analyzed as per instructions given by *Standard Methods for Examination of Water and Wastewater* (APHA 2005). Blank samples were run after every ten

Table 1 | List of industries which are contributing various pollutants in wastewater and groundwater of SIE

Sr.	Type of industry	No.	Pollutant
1.	Aluminum products	6	Metals
2.	Auto parts	8	Metals
3.	Beverage	4	Organics
4.	Chemical manufacturing	6	Organics and metals
5.	Cold storage	2	–
6.	Drugs and pharmaceuticals	4	Organics and metals
7.	Electric goods	6	Organics and metals
8.	Fiber glass industry	4	–
9.	Flour mills	4	Organics
10.	Food products	2	Organics
11.	Glass and glass products	2	–
12.	Knitted textile	2	Dyes and organics
13.	Leather footwears	2	Metals, acids
14.	Light engineering	12	Organics and metals
15.	Marble industry	2	Solids
16.	Packages	6	–
17.	Paints and varnishes	6	Organics
18.	Paper and paper board	2	Lignin
19.	Plastic products	10	Solids
20.	Readymade garments	14	–
21.	Textile processing	8	Organics

samples and spiking of metals was performed as a quality check on an atomic absorption spectrometer (Analyst 800, Perkin Elmer). Samples were run in triplicate and average values were used for statistical analysis. Samples were also sent to the Institute of Chemistry, Punjab University and Pakistan Council for Scientific and Industrial (PCSIR) Laboratories Lahore to reduce the analysis error. An overall ±2.62% error was found.

RESULTS AND DISCUSSION

Drinking and wastewater samples were analyzed, and mean values were used for different statistical tools, i.e., descriptive analysis, Pearson's correlation coefficient, analysis of variance (ANOVA), principal component analysis (PCA), and cluster analysis (CA). These statistical tools were applied for correlation and source of contamination.

Statistical modeling

Descriptive analysis

Descriptive statistics has been computed (Chakrabarty & Sarma 2011). The results of descriptive analysis of all drinking and wastewater samples are shown in Table 2. The mean values of drinking water were also compared with National Drinking Water Quality Standards of Pakistan (NDWQS). The mean value drinking water samples, i.e., pH, turbidity, Ni, Zn, and Cd were within the NDWQS. Bacterial contamination was indicated in drinking water (Table 2), which suggested contamination of drinking water with wastewater. Heavy metals, i.e., Cr, Pb, chlorides, and TDS were beyond the guideline values (Table 2). High concentration of Zn and chloride may be due to fertilizer runoff (Delin & Landon 2002; Valipour 2016). Variations in standard deviation indicated that samples were more representative of the overall study area. Most values of skewness were high, which indicated non-symmetrical distribution of samples (Table 2).

Wastewater samples indicated a very severe picture. Only pH, TDS, chlorides, and Zn were within the limits of National Effluent Quality (NEQ) standards of Pakistan while all other parameters were beyond the limits (Table 2). High values of BOD and COD indicated that the effluents were discharged without any treatment. Skewness has values greater than 1.0 which indicated that the samples were not normal distributed, rather non-symmetrical.

Analysis of variance (ANOVA)

Analysis of variance (ANOVA) is a statistical technique used for the differentiation between two or more means by significance tests (Jalees *et al.* 2016). Mathematically, ANOVA can be measured using the following equation:

$$F = \frac{MST}{MSE} \quad (1)$$

where F is ANOVA coefficient, MST is mean sum of squares due to treatment, and MSE is mean sum of squares due to error.

ANOVA was performed to establish a hypothesis that the parameters (both for drinking and wastewater) have

Table 2 | Descriptive statistics of parameters analyzed for drinking water and wastewater from Sundar Industrial Estate (SIE), Lahore, Pakistan

Parameter	Min	Max	Mean	Std. deviation	Skewness	Guideline values
Drinking water						
pH	7.46	8.1	7.715	0.286	1.0	6.5–8.5
Temp (°C)	22	25	23	1.258	−1.129	–
Turbidity (NTU)	2.25	5.12	3.49	1.19	0.936	<5
Coliform (MPN/100 mL)	32	1,602	1,209.5	785	−2.0	Nil
Fecal coliform (MPN/100 mL)	2	123	47	52.896	1.510	Nil
Total hardness (mg/L $CaCO_3$)	1,480	1,600	1,520	56.569	1.414	<500
Ca^{+2} hardness (mg/L $CaCO_3$)	1,124.8	1,337.6	1,216.2	96.475	0.61	–
Mg^{+2} hardness (mg/L $CaCO_3$)	182.4	355.2	303.8	82.013	−1.850	–
Alkalinity (mg/L $CaCO_3$)	50	662.5	310	256.035	1.031	–
TS (mg/L)	1,708	2,907	2,490	544.287	−1.544	–
TDS (mg/L)	1,024	1,744	1,492	326	−1.53	1,000
TSeS (mg/L)	689	1,263	987.75	234.73	−0.299	–
Chlorides (mg/L)	225	512	349	119.563	0.936	<250
Nickel (ppm)	0.012	0.024	0.019	0.005	−1.056	<0.02
Zinc (ppm)	0.013	0.280	0.088	0.128	1.968	5
Cadmium (ppm)	0.001	0.003	0.002	0.000	−0.060	0.01
Chromium (ppm)	2.117	3.981	3.106	0.944	−0.110	0.05
Lead (ppm)	0.645	1.804	1.206	0.534	0.119	0.05
Wastewater						
pH	6.66	8.89	8.01	1.0	−1.013	6–9
Temp (°C)	12	27	16	7.35	2.0	40 ± 3
BOD (mg/L)	1,619	2,489	2,082	359	−0.4557	80
COD (mg/L)	3,200	5,240	3,830	946.784	1.9124	150
TKN (mg/L)	0.2	173.2	48.2	83.4785	1.9791	–
Total hardness (mg/L $CaCO_3$)	3,120	6,520	4,210	1,557	1.861	–
Ca^{+2} hardness (mg/L $CaCO_3$)	2,464	3,912	2,908	682	1.7865	–
Mg^{+2} hardness (mg/L $CaCO_3$)	624	2,608	1,302	889.1494	1.7500	–
Alkalinity (mg/L $CaCO_3$)	150	700	468.75	232.18	−1.0378	–
TS (mg/L)	223	3,311	2,291	1,410	−1.7455	–
TDS (mg/L)	91	1,329	918	564	−1.7303	3,500
TSeS (mg/L)	133.8	1,984	1,373	845	−1.7470	–
Chlorides (mg/L)	0.0	3.5	1.05	1.6422	1.9348	1,000
Nickel (ppm)	0.0	5.84	1.46	2.92	2.0	1
Zinc (ppm)	0.01	0.05	0.042	0.01	0.01	5
Cadmium (ppm)	0.4316	0.788	0.5651	0.1594	1.2957	0.1
Chromium (ppm)	103	111	108	3.8469	−0.6718	1
Lead (ppm)	97.36	143.16	118.23	19.673	0.5082	0.5

no correlation with each other (null hypothesis). To verify, one-way ANOVA was performed, and results are given in Table 3. For drinking and wastewater sample, F_{crit} is 1.82 and 1.8, respectively, which is smaller than F (36.6 and 18.2) value obtained by ANOVA. This suggested that parameters have some correlation.

Table 3 | Analysis of variance (ANOVA) for drinking water and wastewater

Source of variation	SS	Df	MS	F	F_{crit}
Drinking water					
Between groups	36,739,526	17	2,161,149	32.68	1.82
Within groups	3,571,452	54	66,138		
Wastewater					
Between groups	134,312,814	17	7,900,754	18.25	1.82
Within groups	23,377,023	54	432,907.8		

The null hypothesis and alternative hypothesis were formulated and evaluated using ANOVA.

Pearson correlation coefficient

Pearson correlation coefficient is a helpful statistical approach which measures how strong the relationship is between two or more variables. Mathematically, Pearson correlation coefficient is calculated using the following formula:

$$r = \frac{\sum xy - (\sum x)(\sum y)}{\sqrt{\left\{N\sum x^2 - (\sum x)^2\right\} nx \left\{N\sum y^2 - (\sum y)^2\right\}}} \quad (2)$$

where N is number of pair of scores, while x and y are two variables.

Pearson correlation coefficient was performed for drinking and wastewater to identify the extent of correlation among parameters. The results of Pearson correlation coefficient are given in Table 4. In the case of drinking water, pH showed strong correlation with temperature, total solids, total dissolved solids, total hardness, alkalinity, and zinc. Temperature and turbidity showed strong correlation with coliform and fecal coliform, which indicated that wastewater is contaminating the drinking water. Heavy metals showed moderate to strong correlation (0.3–0.9) among drinking water samples. In the case of wastewater samples, pH showed moderate to strong correlation (0.3–0.7) with BOD and COD, respectively (Table 4). BOD and COD both showed moderate to strong correlation with heavy metals, which suggested that heavy metals were contributed from industrial effluents of SIE (Table 4). Heavy metals showed strong correlation among them, which indicated the same source of these metals. Chlorides showed strong correlation (0.7–0.9) with heavy metals, which suggested that the salts of metals were present in industrial effluents. The common industrial salts are nitrates, chlorides, and sulphate of metals.

Principal component analysis (PCA)

The principal component analysis (PCA) is a tool which is based on an imaginary Eigen value. In the present study, all Eigen values which were less than 1 were ignored. The components having Eigen value >1 are grouped based on the same source (Jalees *et al.* 2016). PCA using the rotation method of varimax and Kaiser normalization was performed on drinking water and wastewater samples and the results are shown in Table 5. The PCA for drinking water gave three components, named as PC 1, PC 2, and PC 3, which explained a total of >98% of total variance. PC 1 explained 56.69% of the total variance, PC 2 explained 26.5% while PC 3 explained 16.78% (Table 5). PC 1 expressed highest loading for pH, temperature, coliform, total hardness, alkalinity, TS, TDS, TSeS, and zinc. This high loading reflected seepage to groundwater aquifer from sewage effluent discharges, urban runoff, industrial waste discharges, and contamination from refuse leachate to the ultimate problem (WHO 1984; Jalees *et al.* 2016). SIE has various industries which contribute these pollutants (Table 1). Moreover, dissolution of salt deposits in the aquifer can increase heavy metal levels, and waters in the areas of Paleozoic and Mesozoic sedimentary rock have higher TDS levels, ranging from as little as 195 to 1,100 mg/L (WHO 1984). PC 2 showed the highest loading for turbidity, Cl^{-1}, Cd, Cr, and Pb which indicated dissolution of rocks and minerals in the aquifer or anthropogenic activities (WHO 1984). PC 3 showed maximum loading for calcium hardness and Ni reflected contamination from the seepage of industrial emissions and tanneries' wastewater (Table 5) (WHO 1984; Jalees *et al.* 2016).

For wastewater, three components were identified as PC 1, PC 2, and PC 3 with a combined total of variance of >98%. PC 1 explained 63.76%, PC 2 explained 25.32, and PC 3 explained 10.9% of variance. PC 1 gave maximum loading for pH, temperature, hardness, and heavy metals which indicated metal precipitation due to pH. PC 2 indicated that BOD and solids had high loading values which

Table 4 | Pearson correlation coefficient analysis for drinking and wastewater parameters to find extent of correlation among parameters

	pH	Temp	Turb.	Coli form	Fecal form	Total hardness	Ca hardness	Mg hardness	Alk	TS	TDS	TSeS	Cl^-	Ni	Zn	Cd	Cr	Pb
Drinking water parameters																		
pH	1.000																	
Temp	−0.671	1.000																
Turbidity	−0.020	−0.020	1.000															
Coliform	−0.897	0.927	0.056	1.000														
Fecal form	−0.652	0.336	0.764	0.567	1.000													
Total hardness	0.988*	−0.749	−0.116	−0.943	−0.704	1.000												
Ca^{+2} hardness	0.597	0.159	−0.320	−0.220	−0.697	0.530	1.000											
Mg^{+2} Hardness	−0.021	−0.704	0.296	−0.392	0.335	0.067	−0.811	1.000										
Alkalinity	0.890	−0.760	−0.406	−0.918	−0.847	0.946	0.475	0.094	1.000									
TS	−0.795	0.964*	−0.185	0.958*	0.308	−0.835	0.011	−0.590	−0.766	1.000								
TDS	−0.793	0.964*	−0.189	0.956*	0.304	−0.833	0.015	−0.592	−0.763	1.000**	1.000							
TSeS	−0.534	0.985*	−0.049	0.848	0.216	−0.624	0.325	−0.812	−0.653	0.919	0.920	1.000						
Chlorides	−0.020	−0.020	1.000**	0.056	0.764	−0.116	−0.320	0.296	−0.406	−0.185	−0.189	−0.049	1.000					
Ni	0.507	0.052	−0.714	−0.261	−0.911	0.508	0.888	−0.694	0.606	0.023	0.028	0.191	−0.714	1.000				
Zn	0.918	−0.901	−0.102	−0.997	−0.621	0.962*	0.285	0.327	0.942	−0.935	−0.933	−0.813	−0.102	0.331	1.000			
Cd	−0.352	0.774	0.566	0.668	0.572	−0.493	0.173	−0.544	−0.709	0.598	0.596	0.788	0.566	−0.199	−0.662	1.000		
Cr	−0.095	0.595	0.660	0.440	0.479	−0.248	0.294	−0.517	−0.518	0.371	0.370	0.642	0.660	−0.140	−0.433	0.962*	1.000	
Pb	0.527	−0.752	−0.644	−0.747	−0.761	0.651	0.090	0.343	0.850	−0.617	−0.614	−0.722	−0.644	0.431	0.757	−0.965	−0.890	1.000

Wastewater parameters	pH	Temp	BOD	COD	TKN	Total hardness	Ca hardness	Mg hardness	Alkalinity	TS	TDS	TSeS	Chloride	Ni	Cd	Cr	Pb
pH	1.0000																
Temp	−0.8758	1.0000															
BOD	−0.3878	0.7808	1.0000														
COD	0.6668	−0.3852	0.0433	1.0000													
TKN	0.5340	−0.2255	0.1697	0.985*	1.0000												
Total hardness	−0.9234	0.989*	0.6883	−0.4060	−0.2444	1.0000											
Ca^{+2} hardness	−0.965*	0.972*	0.6138	−0.5367	−0.3851	0.989*	1.0000										
Mg^{+2} hardness	−0.8772	0.987*	0.7350	−0.2994	−0.1327	0.993**	.965*	1.0000									
Alkalinity	−0.7736	0.6154	0.2569	−0.954*	−0.8999	0.6103	0.7139	0.5215	1.0000								
TS	0.0567	0.3700	0.8314	0.0909	0.1322	0.2278	0.1738	0.2658	0.1491	1.0000							
TDS	0.0530	0.3719	0.8310	0.0840	0.1253	0.2297	0.1767	0.2669	0.1555	1.000**	1.0000						
TSeS	0.0558	0.3710	0.8321	0.0909	0.1324	0.2288	0.1748	0.2668	0.1493	1.000**	1.000**	1.0000					
Chlorides	−0.9141	0.986*	0.7212	−0.5313	−0.3830	0.976*	0.983*	0.955*	0.7377	0.3456	0.3486	0.3465	1.0000				
Ni	−0.8969	0.998**	0.7552	−0.4436	−0.2875	0.989*	0.981*	0.979*	0.6640	0.3530	0.3553	0.3540	0.995**	1.0000			
Cd	−0.996**	0.9132	0.4678	−0.6487	−0.5107	0.9484	0.983*	0.9071	0.7791	0.0331	0.0367	0.0340	0.9472	0.9320	1.0000		
Cr	0.986*	−0.8221	−0.2867	0.6057	0.4740	−0.8912	−0.9290	−0.8485	−0.6866	0.2000	0.1969	0.1991	−0.8503	−0.8405	−0.970*	1.0000	
Pb	0.9215	−0.6643	−0.1174	0.8927	0.8118	−0.7178	−0.8123	−0.6343	−0.9072	0.1855	0.1799	0.1849	−0.7612	−0.7072	−0.9005	0.8988	1.0000

Table 5 | Principal component analysis (PCA) for source identification

	1	2	3
Drinking water			
Total variance (%)	56.69	26.52	16.78
pH	−0.8869	0.0266	0.4612
Temp	0.9235	0.2283	0.3082
Turbidity	−0.1329	0.9263	−0.3526
Coliform	0.9822	0.1763	−0.0653
Fecal coliform	0.4198	0.6402	−0.6434
Total hardness	−0.9146	−0.1085	0.3897
Ca^{+2} hardness	−0.1594	0.0075	0.9872
Mg^{+2} hardness	−0.4432	−0.0836	−0.8925
Alkalinity	−0.8350	−0.4251	0.3495
TS	0.9853	0.0071	0.1707
TDS	0.9847	0.0040	0.1744
TSeS	0.8506	0.2463	0.4646
Chlorides	−0.1329	0.9263	−0.3526
Ni	−0.1262	−0.4530	0.8825
Zn	−0.9709	−0.1987	0.1338
Cd	0.5560	0.7896	0.2596
Cr	0.3112	0.8870	0.3411
Pb	−0.6197	−0.7848	−0.0027
Wastewater			
Total variance (%)	63.76	25.32	10.9
pH	−0.899	0.104	0.426
Temp	0.938	0.312	−0.154
BOD	0.596	0.788	0.155
COD	−0.280	0.078	0.957
TKN	−0.114	0.107	0.988
Total hardness	0.974	0.167	−0.153
Ca^{+2} hardness	0.949	0.119	−0.293
Mg^{+2} hardness	0.979	0.201	−0.043
Alkalinity	0.464	0.147	−0.874
TS	0.069	0.997	0.034
TDS	0.069	0.997	0.027
TSeS	0.070	0.997	0.034
Chlorides	0.902	0.295	−0.316
Ni	0.930	0.297	−0.216
Cd	0.912	−0.016	−0.410
Cr	−0.903	0.251	0.349
Pb	−0.658	0.207	0.724

suggested that solids in untreated wastewater were responsible for high BOD contents (Jalees *et al.* 2016). PC 3 gave high loading for COD, TKN, and alkalinity, which suggested that these contaminations originated from the same source.

Cluster analysis (CA)

A tree diagram, which shows the agglomerative hierarchical clustering algorithms available in the data, is called a dendrogram (Hintze 1995). This diagram is used for the extent of correlation among the parameters. The CA dendrogram (single linkage) was performed on average parameter values of drinking water and wastewater samples from all locations, as shown in Figure 2. For drinking water, the elucidation distance of turbidity, chlorides, and hardness comprises group G_1; elucidation distance of Ni, Zn, Pb, Cr, pH, temperature, and fecal coliform comprises G_2 while elucidation distance of solids and hardness comprises G_3 (third group). All these groups have strong correlation within groups as indicated by the small elucidation distance (Figure 2), but among the groups, these parameters showed long elucidation distance which indicated the different sources of these pollutants in the drinking water (Jalees *et al.* 2016). In the wastewater sample, the elucidation distance of heavy metals, i.e., Ni, Cd, Cr, Pb along with chlorides, pH, temperature, TKN showed a similar origin and all fall under one group with strong correlation among group members. Other than these, all wastewater parameters have a long elucidation distance which indicated the multiple sources of pollutants present in the wastewater (Jalees *et al.* 2016).

Risk assessment

Lifetime average daily dose (LADD)

The probability of health risk at sites where contaminated water is used for drinking purposes is very high. There is the potential that the population drinking this contaminated water may suffer from cancer or life-threatening diseases. Such health risk is calculated by measuring the concentration of contaminants in drinking water. Using contaminations' concentration, the lifetime average daily

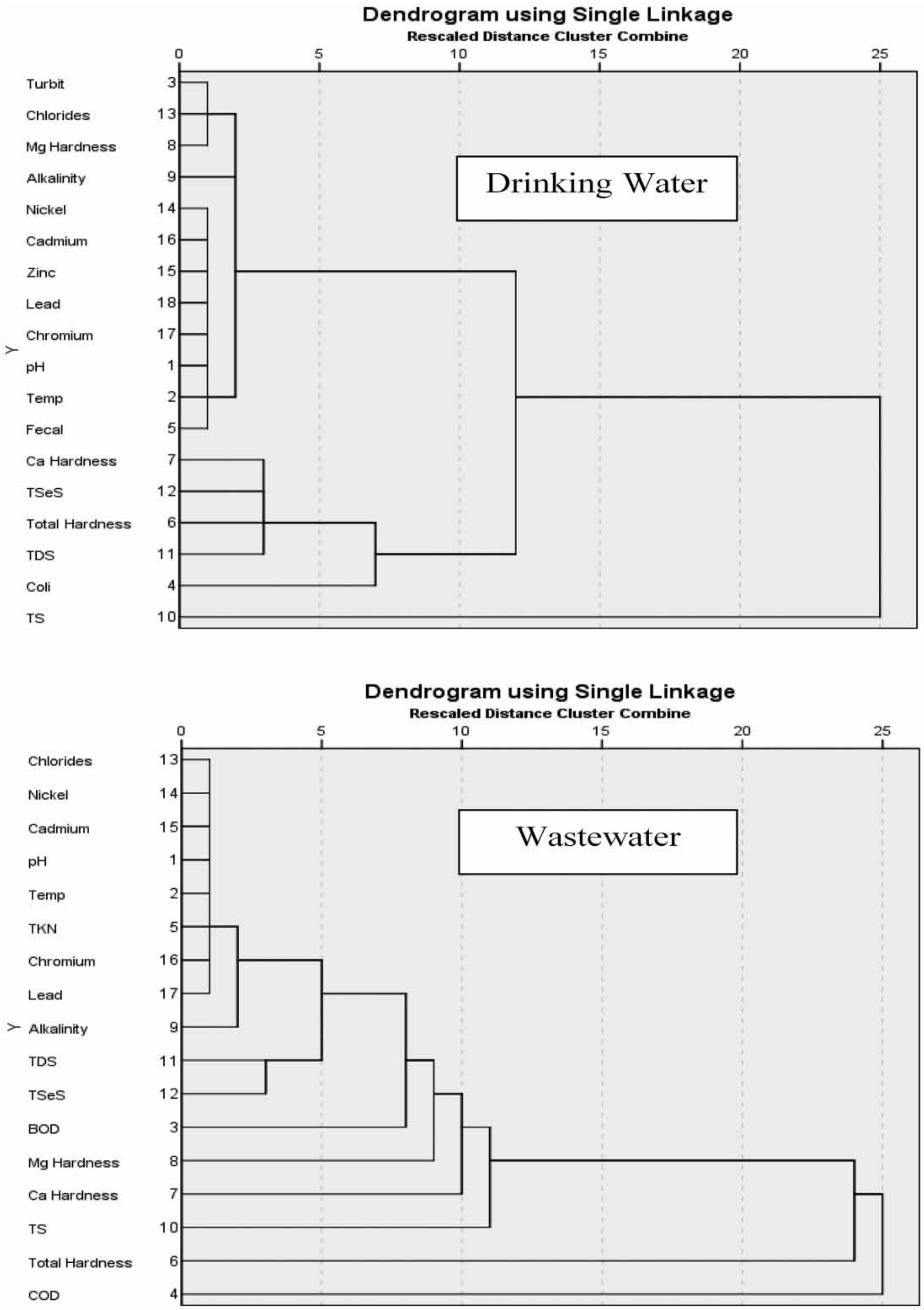


Figure 2 | Dendrogram of parameters for drinking and wastewater samples from Sunder Industrial Estate for source identification.

dose (LADD) (EPA 2004) is calculated. The formula to calculate *LADD* is

$$LADD = \frac{CxCFxIRxEFxED}{BWxAT} \quad (3)$$

where *C* is concentration of contamination in drinking water (mg/L); *CF* is conversion factor for 1,000 mL/L; *IR* is intake rate (L/day); *EF* is exposure frequency (days/year); *ED* is exposure duration (years); *BW* is body weight for population of interest (kg); and, *AT* is average time (days).

The values for *LADD* for heavy metals, i.e., Ni, Pb, Cr, Cd, and Zn, were calculated using Equation (3) (EPA 2004). The Environmental Protection Agency of the United States (US EPA) developed this relation in 2004. In locations where drinking water is contaminated, the potential may exist for uptake via ingestion (drinking). This may result in exposure of toxic pollutants among local populations in the contaminated area. Receptors could include families living nearby or societies. Exposure via intake/drinking of contaminated water depends upon the concentration of pollutant intake but also the rate at which the water is used, and the frequency and duration of exposure.

The calculated values in drinking water are given in Table 6. The values of Ni were in the range of 1.6×10^{-4} to 2.1×10^{-4}; for Pb, values ranged from 7.6×10^{-3} to 1.3×10^{-2}; for Cr, the values were 1.8×10^{-3} to 3.4×10^{-3}; for Cd, the values were 6.9×10^{-7} to 3.4×10^{-6}; and for Zn, values were 2.1×10^{-4} to 3.1×10^{-4} (Table 6). Based on the average *LADD* values, the metals showed the following trend Cr > Pb > Zn > Ni > Cd.

Non-carcinogenic assessment

For standard risk assessment, it is assumed that there is a threshold level of exposure to chemicals. If the exposure is less than the threshold then no adverse effects will be observed (EPA 2004). The default approaches used in this study to assess the potential for health effects was based on the US EPA approach, i.e., comparing an estimate of ingested exposure to an R_fD for oral exposures. An R_fD is a daily oral intake rate that is estimated to pose no appreciable risk of adverse health effects, even to sensitive populations, over a 70-year lifetime (EPA 2004). To calculate the non-cancer risk, *HQ* is determined using *LADD* and R_fD. The threshold value in this study is 1 (EPA 2004). The potential for non-carcinogenic effects is determined

Table 6 | Risk assessment calculation showing lifetime average daily dose (*LADD*), hazard quotient (*HQ*) and cancer risk for heavy metals present in drinking water of SIE

	Ni	Pb	Cr	Cd	Zn
LADD					
Location 1	2.1×10^{-4}	7.6×10^{-3}	3.4×10^{-2}	3.4×10^{-6}	3.1×10^{-4}
Location 2	1.6×10^{-4}	13×10^{-3}	1.8×10^{-2}	1.1×10^{-6}	2.1×10^{-4}
Location 3	1.8×10^{-4}	8.6×10^{-3}	2.1×10^{-2}	0.69×10^{-6}	24×10^{-4}
Location 4	2.1×10^{-4}	7.6×10^{-3}	3.4×10^{-2}	3.4×10^{-6}	3.1×10^{-4}
R_fD	0.02	0.004	1.5	0.001	0.3
Hazard quotient					
Location 1	5.1×10^{-3}	1.4	2.2×10^{-2}	3.3×10^{-3}	3.7×10^{-4}
Location 2	10×10^{-3}	1.9	2.3×10^{-2}	3.4×10^{-3}	10×10^{-4}
Location 3	8.2×10^{-3}	3.2	1.2×10^{-2}	1.1×10^{-3}	6.9×10^{-4}
Location 4	9.0×10^{-3}	2.1	1.4×10^{-2}	6.9×10^{-3}	80×10^{-4}
HI	33×10^{-3}	8.6	7.1×10^{-2}	8.5×10^{-3}	100×10^{-4}
Total hazard index (*HI*)			**8.7**		
Cancer risk					
Location 1	9.4×10^{-5}	4.7×0^{-5}	1.6×10^{-2}	4.9×10^{-6}	–
Location 2	19×10^{-5}	6.5×10^{-5}	1.7×10^{-2}	5.2×10^{-6}	–
Location 3	15×10^{-5}	11×10^{-5}	91×10^{-2}	1.7×10^{-6}	–
Location 4	16×10^{-5}	7.3×10^{-5}	1.1×10^{-2}	1.0×10^{-6}	–
Slope factor	0.91	0.0085	0.5	1.5	–
Sum	59×10^{-5}	29×10^{-5}	93×10^{-2}	13×10^{-6}	–
Total cancer risk			93×10^{-2}		

using the formula (EPA 2004):

$$HQ = \frac{LADD}{R_f D} \quad (4)$$

where R_fD is standard reference daily dose for each metal. The values of R_fD (Table 6) are obtained from Integrated Risk Information System (EPA 2019). Equation (4) is widely used for the assessment of non-cancer effects, e.g., fluoride in water (Satou *et al.* 2020), personal care products in coastal wetland (Sadutto *et al.* 2020), heavy metals risk assessment (Özden *et al.* 2020; Yahaya *et al.* 2020), phthalates exposure in children (Søeborg *et al.* 2012), consumption of vegetables (Khan *et al.* 2009). The *HQ* values calculated using Equation (4) are given in Table 6. The *HQ* of Ni is 5.1×10^{-3} to 1×10^{-2}; for Pb, 1.4 to 3.2; for Cd, 6.9×10^{-4} to 3.4×10^{-3}; for Cr, 1.2×10^{-2} to 2.3×10^{-2}; and for Zn, 3.7×10^{-4} to 8×10^{-3} (Table 6). The total sum of *HQ* of every metal was in the order of Pb > Cr > Ni > Zn > Cd. The values of *HQ* (Table 5) for heavy metals are less than 1 except in Pb, which indicated that there is a risk of non-carcinogenic effect in the study area due to Pb while for other metals the non-carcinogenic effects are within the permissible limit (EPA 2019). As the people in the study area are exposed to multiple heavy metals it is necessary to determine total non-carcinogenic hazard through all heavy metals (EPA 2019), i.e., *HI*. The *HI* is determined through the following equation (EPA 2019):

$$HI = \sum_i HQ_i \quad (5)$$

The value of *HI* (Table 6) is 8.7 which clearly indicates that there is a probability of non-carcinogenic hazard among people in the study area due to consumption of contaminated drinking water.

Cancer risk assessment

The human health risk models for carcinogenic assessment developed by US EPA, have proved successful and been adopted worldwide. Currently, there is no agreed limit for acceptable maximum carcinogenic levels in Pakistan, therefore, the US EPA model was adopted in this study. The health risk assessment was divided into four steps: (i) hazard identification; (ii) dose response assessment; (iii) exposure assessment; (iv) risk characterization (USEPA 1997; EPA 2019). This multiphase and multicomponent risk assessment model (Equations (6) and (7)) developed by US EPA was used to evaluate the heavy metal pollution hazard in drinking water (EPA 2019). The calculations for the lifetime average daily dose of contaminants and the detailed explanation for all the parameters are provided in Equations (3), (6) and (7) (USEPA 1997; Liu *et al.* 2013; EPA 2019). Cancer risk (*CR*) assessment is determined using *LADD* values and slope factor (*SP*) by the following equation (Liu *et al.* 2013):

$$CR = LADDxSF \quad (6)$$

where *SF* is an upper bound (95% percentile) cancer risk due to lifetime exposure of heavy metals through drinking water (USEPA 1997). Equation (6) is widely used for the estimation of cancer risk assessment, e.g., heavy metals in vegetables (Liu *et al.* 2013; Sultana *et al.* 2017), heavy metals in street dust (Zheng *et al.* 2010), heavy metals in mining (Lim *et al.* 2008). The *CR* values for heavy metals are given in Table 6. The *CR* value for Ni was 9.5×10^{-5} to 1.9×10^{-4}; for Pb, 4.7×10^{-5} to 1×10^{-4}; for Cr, 9.1×10^{-3} to 1.7×10^{-2}; and for Cd, 1×10^{-6} to 4.9×10^{-6} (Table 6). As all values of *CR* for heavy metals are above 1×10^{-6} (USEPA 1997) this clearly indicates that drinking water in SIE poses a serious cancer risk for the surrounding population. Based on the sum of *CR* of each metal from all locations, the order of CR is Cr > Ni > Pb > Cd. As the population of the study area is exposed to more than one heavy metal, the total risk through drinking water is calculated using the following equation (EPA 2004):

$$Total\ Cancer\ Risk_T = \sum_i^n Cancer\ Risk_i \quad (7)$$

where *n* is number of metals and *i* represents the respective metal. The total cancer risk value for the study area was 93×10^{-2} which is quite high as compared to the permissible limit, i.e., 1×10^{-6}. This suggested that 93 people out of every 100 are at risk for carcinogenic effect.

CONCLUSIONS

In this study, drinking water and wastewater samples were collected from Sundar Industrial Estate for source, correlation, and health risk assessment. The analysis of drinking water indicated high bacterial and heavy metals' contamination which is beyond the NDWQS of Pakistan. High BOD and COD values of wastewater indicated the presence of untreated effluent discharge. Descriptive statistics indicated non-symmetrical distribution of parameters except in the case of heavy metals for both drinking and wastewater. ANOVA and Pearson correlation indicated moderate to strong correlation among physical and chemical parameters of drinking and wastewater analysis. PCA and CA indicated the seepage of wastewater, contamination from refuse leachate, and untreated effluent discharges as the main sources of contamination for drinking water. Lifetime average daily dose (*LADD*) for heavy metals is in the order of Cr > Pb > Zn > Ni > Cd. The *HQ* and *HI* indicated the probability of non-carcinogenic risk as values are greater than 1. Cancer risk assessment and total risk assessment showed that 93 people from every 100 of population may suffer from cancer due to drinking water contaminated by industrial and anthropogenic activities. Based on the results of the current study, it is recommended that the government should implement the regulation of wastewater treatment strictly. In addition to this, a detailed study of other industrial estates situated in different cities of Pakistan should also be undertaken with reference to health risk assessment so that an overall country picture can be seen, and proper remedies can be implemented.

ACKNOWLEDGEMENTS

The authors are grateful to Institute of Chemistry, Punjab University, Lahore and Pakistan Council of Scientific and Industrial Research Laboratories, Lahore for providing analytical facilities.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

APHA 2005 *Standard Methods for the Examination of Water and Wastewater*. American Public Health Association (APHA), Washington DC, USA.

Azizullah, A., Khattak, M. N. K., Richter, P. & Häder, D.-P. 2011 Water pollution in Pakistan and its impact on public health – a review. *Environment International* **37** (2), 479–497.

Baig, S., Xu, X. & Khan, R. 2012a Microbial water quality risks to public health: potable water assessment for a flood-affected town in northern Pakistan. *Rural Remote Health* **12** (3), 2196.

Baig, S. A., Akhter, N. A., Ashfaq, M., Asi, M. R. & Ashfaq, U. 2012b Imidacloprid residues in vegetables, soil and water in the southern Punjab, Pakistan. *Journal of Agricultural Technology* **8** (3), 903–916.

Chakrabarty, S. & Sarma, H. P. 2011 A statistical approach to multivariate analysis of drinking water quality in Kamrup district, Assam, India. *Archives of Applied Science Research* **3** (5), 258–264.

Delin, G. N. & Landon, M. K. 2002 Effects of surface run-off on the transport of agricultural chemicals to ground water in a sandplain setting. *Science of the Total Environment* **295** (1–3), 143–155.

EPA 2004 *Examples of Exposure Scenarios*. National Center for Environmental Assessment, Environmental Protection Agency, Washington DC, USA, p. 6.

EPA 2019 *Integerated Risk Information System*. Environmental Protection Agency, Washington DC, USA.

Hintze, J. 1995 *NCSS Statistical Software*. Kayville, UT, USA.

Jalees, M. I., Aslam, A., Fatima, R., Khalid, I. & Hasan, B. 2016 Statistical modeling of groundwater quality for source and ionic relationships: a case study for drinking water quality. *Procceding of the Pakistan Academy of Sciences: B. Life and Environmental Sciences* **53** (1), 11.

Khan, S., Farooq, R., Shahbaz, S., Khan, M. A. & Sadique, M. 2009 Health risk assessment of heavy metals for population via consumption of vegetables. *World Applied Sciences Journal* **6** (12), 1602–1606.

Khan, K., Khan, H., Lu, Y., Ihsanullah, I., Nawab, J., Khan, S., Shah, N. S., Shamshad, I. & Maryam, A. 2014 Evaluation of toxicological risk of foodstuffs contaminated with heavy metals in Swat, Pakistan. *Ecotoxicology and Environmental Safety* **108**, 224–232.

Lim, H.-S., Lee, J.-S., Chon, H.-T. & Sager, M. 2008 Heavy metal contamination and health risk assessment in the vicinity of the abandoned Songcheon Au–Ag mine in Korea. *Journal of Geochemical Exploration* **96** (2–3), 223–230.

Liu, X., Song, Q., Tang, Y., Li, W., Xu, J., Wu, J., Wang, F. & Brookes, P. C. 2013 Human health risk assessment of heavy metals in soil–vegetable system: a multi-medium analysis. *Science of the Total Environment* **463**, 530–540.

Özden, Ö., Can Tunçelli, İ., Oray, I. K., Kaplan, M., Parıldar, S. & Erkan, N. 2020 Heavy metal risk assessment of European eels (*Anguilla anguilla*, Linnaeus, 1758) from the Asi (Orontes)

River, Turkey. *Journal of Applied Ichthyology* **36** (6), 912–917.

PIEDMC 2006 *Environmental Impact Assessment of Sundar Industrial Estate*. Punjab Inudstrial Estate Development and Management Company, Lahore, Pakistan, p. 204.

Raza, M., Hussain, F., Lee, J.-Y., Shakoor, M. B. & Kwon, K. D. 2017 Groundwater status in Pakistan: a review of contamination, health risks, and potential needs. *Critical Reviews in Environmental Science and Technology* **47** (18), 1713–1762.

Rehman, W., Zeb, A., Noor, N. & Nawaz, M. 2008 Heavy metal pollution assessment in various industries of Pakistan. *Environmental Geology* **55** (2), 353–358.

Sadutto, D., Andreu, V., Ilo, T., Akkanen, J. & Picó, Y. 2020 Pharmaceuticals and personal care products in a Mediterranean coastal wetland: impact of anthropogenic and spatial factors and environmental risk assessment. *Environmental Pollution* **271**, 116353.

Saeed, T. U. & Attaullah, H. 2014 Impact of extreme floods on groundwater quality (in Pakistan). *British Journal of Environment and Climate Change* **4** (1), 133–151.

Saiqa, I., Kiran, A., Lubna, B. & Muhammad, A. 2016 *Water Quality Status of Major Cities of Pakistan*. Pakistan Council for Research in Water Resources, Islamabad, Pakistan, p. 110.

Satou, R., Aikawa, H. & Sugihara, N. 2020 Risk assessment by seasonal variation of well water fluoride in Japan. *French-Ukrainian Journal of Chemistry* **8** (2), 104–112.

Søeborg, T., Frederiksen, H. & Andersson, A.-M. 2012 Cumulative risk assessment of phthalate exposure of Danish children and adolescents using the hazard index approach. *International Journal of Andrology* **35** (3), 245–252.

Sultana, M. S., Islam, M. S., Saha, R. & Al-Mansur, M. 2017 Health risk assessment for carcinogenic and non-carcinogenic heavy metal exposures from vegetables and fruits of Bangladesh. *Cogent Environmental Science* **3** (1), 1291107.

USEPA 1997 *Exposure Factors Handbook*. Office of Research and Development, Washington DC, USA.

Valipour, M. 2016 How do different factors impact agricultural water management? *Open Agriculture* **1** (1), 89–111 (open-issue).

Valipour, M. 2017 Global experience on irrigation management under different scenarios. *Journal of Water and Land Development* **32** (1), 95–102.

WHO 1984 *Guidelines for Drinking Water Quality, Vol. 2*. World Health Organization, Geneva, Switzerland.

Yahaya, T. O., Oladele, E. O., Fatodu, I. A., Abdulazeez, A. & Yeldu, Y. I. 2020 The concentration and health risk assessment of heavy metals and microorganisms in the groundwater of Lagos, Southwest Nigeria. *Journal of Advances in Environmental Health Research* **8** (3), 234–242.

Zheng, N., Liu, J., Wang, Q. & Liang, Z. 2010 Health risk assessment of heavy metal exposure to street dust in the zinc smelting district, Northeast of China. *Science of the Total Environment* **408** (4), 726–733.

First received 30 October 2020; accepted in revised form 8 January 2021. Available online 15 February 2021