For countries such as Bangladesh with a significant groundwater arsenic problem, there is an urgent need for the arsenic-contaminated wells to be identified as soon as possible and for appropriate action to be taken. This will involve the testing of a large number of wells, potentially up to 11 million in Bangladesh alone. Field-test kits offer the only practical way forward in the timescale required. The classic field method for detecting arsenic (the 'Gutzeit' method) is based on the reaction of arsine gas with mercuric bromide and remains the best practical approach. It can in principle achieve a detection limit of about 10 µg l^{-1} by visual comparison of the coloured stain against a colour calibration chart. A more objective result can be achieved when the colour is measured by an electronic instrument. Attention has to be paid to interferences mainly from hydrogen sulfide. Due to analytical errors, both from the field-test kits and from laboratory analysis, some misclassification of wells is inevitable, even under ideal conditions. The extent of misclassification depends on the magnitude of the errors of analysis and the frequency distribution of arsenic observed, but is in principle predictable before an extensive survey is undertaken. For a country with an arsenic distribution similar to that of Bangladesh, providing care is taken to avoid sources of bias during testing, modern field-test kits should be able to reduce this misclassification to under 5% overall.

Keywords

---

1. Introduction

The high concentrations of arsenic now being found in groundwater in many parts of the world [1] pose an important analytical challenge because of the large number of wells that must be tested. This is particularly true in Bangladesh. Most rural Bangladeshis now derive their drinking water from tubewells drilled in the alluvial and deltaic deposits that make up a large part of the country. It is estimated that there are some 6–11 million shallow tubewells in Bangladesh and that according to the DPHE/BGS 1998/99 survey of 61 out of the 64 districts in Bangladesh, approximately 27% of the shallow wells (<150 m depth) sampled are contaminated with arsenic above the Bangladesh drinking water standard of 50 µg l^{-1}[2]. The median total As concentration in these wells was 6 µg l^{-1} (Fig. 1). The greatest concentration of affected wells is in the south and east of Bangladesh, and in the 12 worst-affected districts (Chandpur, Madaripur, Munshiganj, Gopalganj, Lakshmipur, Noakhali, Bagerhat, Shariatpur, Comilla, Faridpur, Satkhira and Meherpur), the median arsenic concentration is 135 µg l^{-1} and some 76% of shallow wells exceed the Bangladesh standard for arsenic. 60% or more of the wells in each of these worst-affected districts contained arsenic concentrations exceeding 50 µg l^{-1} and occasionally concentrations exceeded 1000 µg l^{-1}. More than 10 000 arsenic-affected patients have already been diagnosed in Bangladesh, and there is a fear that this number under-represents the true number of people affected and that the number could increase rapidly if the contaminated wells continue to be used for drinking water [3].
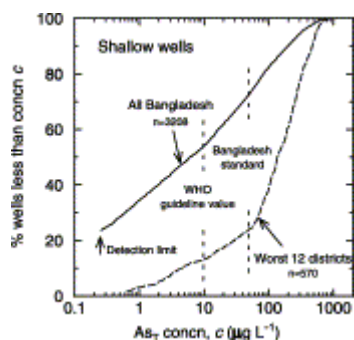
Fig. 1. Cumulative frequency distribution of arsenic in shallow tubewells in Bangladesh for both the whole of Bangladesh (excluding the Chittagong Hill Tracts) and for the worst 12 arsenic-affected districts. The WHO guideline value (10 µg l^{-1}) and Bangladesh standard for arsenic in drinking water (50 µg l^{-1}) are indicated.

While there is clearly a distinct regional pattern in the distribution of arsenic concentrations in Bangladesh, at the village-scale there is also a large amount of well-to-well variation and ultimately each well will have to be tested for arsenic. Some 10% of the tubewells have so far (January 2002) been tested and plans are in place to analyse the remainder as rapidly as possible. Similar screening programmes also need to be undertaken in other 'at risk' areas of the world, including many of the delta regions of Asia [1] where As testing facilities are poor. Cheap but robust methods of analysis are needed. Training and quality control are also important particularly when the majority of wells are privately owned and testing may be carried out by the private sector. Here we review some of the methods of testing drinking water for arsenic in the field and, using Bangladesh as an example, analyse in some detail the impact that random analytical errors can have on the decision as to whether to classify a well as 'safe' or 'not safe'.

2. Methods suitable for field analysis of arsenic

2.1. Arsine gas generation

Most of the analytical methods for arsenic, including field methods, take advantage of the formation of volatile arsine (AsH$_3$) gas to separate the arsenic from other possible interferences in the sample matrix [4].[1] The classical approach for generating arsine gas is by reduction using zinc and hydrochloric acid to generate 'nascent' hydrogen. Since the determination of arsenic in drinking water involves trace analysis in the micrograms per litre range, it is very important to use high-purity chemicals. As arsenic is always associated with zinc ores, this is a particular problem and the refinery process to obtain low-As zinc is costly. Zinc of at least 99.9% purity is required and it should have an As content of <0.1 mg kg^{-1}. Many of the sources of zinc used for the early arsenic testing in Bangladesh and elsewhere have had relatively high degrees of arsenic contamination. This leads to a relatively high detection limit. The need for low-arsenic zinc is of increasing importance as the desire to measure arsenic concentrations at 10 µg l^{-1} (the WHO guideline value for arsenic in drinking water) and below increases. The quality of the Zn and other chemicals is best assessed from regular 'blank' determinations.

Another factor in the performance of the test kits is the grain size of the zinc. This affects the kinetics of hydrogen production. The generation of the arsenic hydride and the formation of

excess hydrogen gas as a carrier gas have to continue for long enough to expel all of the arsenic from the sample, yet be sufficiently rapid to enable a reasonable sample throughput. Zinc granules (10–100 mesh) are preferable to powder in this respect. Under the prevailing operating conditions of most field-test kits, the reduced form of arsenic (arsenite or As(III)) generates arsine gas most rapidly, with the response of arsenate (As(V)) being somewhat slower. Details of the factors controlling the kinetics are unclear, but matrix effects appear to be significant. For example, the reduction of As(V) appears to be catalysed by Cl and/or heavy metals. Therefore, the As(V) species are often reduced to As(III) species in a separate pre-reduction step. This is usually carried out by adding a few drops of concentrated potassium iodide and stannous chloride solution. This pre-reduction step is not very rapid and so increases the time required for a single determination. Most field-test kits do not include a pre-reduction step but rely on the reduction of As(V) directly to AsH_3 under strongly acidic conditions.

A practical weakness of this classical approach is the use of rather concentrated hydrochloric acid, often 6 N HCl. This not only raises problems during transport, but it is also rather dangerous for use by technicians without formal chemistry training. An important step introduced with the Arsenator[2] has therefore been to replace this strong acid with sulfamic acid, NH_2SO_3H, in solid form. Sulfamic acid has also been adopted in the Hach arsenic test kit (Hach Company, Loveland, CO, see [5]). Since the acidity of sulfamic acid is less than that of HCl, the time taken for a single determination is greater, particularly for samples having a naturally high alkalinity (which is a characteristic of many arsenic-rich groundwaters).

The most reliable way to carry out hydride generation (HG) in the field is to replace the zinc with the strongly reducing sodium borohydride, $NaBH_4$. This reagent is widely used in the laboratory when using HG with atomic absorption spectrophotometry (HG–AAS) or inductively coupled plasma–atomic emission spectrometry (HG–ICP–AES). Arsenic contamination of sodium borohydride is also less than with solid zinc, although the extent of contamination varies with supplier. When combined with a solid acid, sodium borohydride has the advantage that the analysis can be carried out without any liquids other than the sample. Excess borohydride can be used to generate hydrogen gas which then acts as a carrier for arsine. A further important advantage of such a combination is the fact that no additional steps have to be taken to ensure the reduction of As(V) to As(III). Reduction of As(V) to arsine (AsH_3) with solid sodium borohydride is rapid for reasons that are not entirely clear. This reduces the time for an analysis to a few minutes. Although arsine gas is very toxic to humans, it should not pose a serious hazard in the practical analysis. The amounts of arsine generated during analysis of groundwater are normally small, and with an efficient generator and absorption system, most of the arsine gas reacts with the colour-forming reagent and is therefore removed from the environment. However, it is of course wise to work in a well-ventilated environment in order to avoid any possible harm [6].

While arsenic species can be detected using various electrochemical methods, it is unlikely that such methods could ever be made robust enough and cheap enough for incorporation into routine field-test kits. Electrodes are notoriously fickle, and usually expensive and difficult to replace. Other methods involving biosensors are under development but have not yet been demonstrated to be specific enough for the routine testing of arsenic in groundwater.

2.2. Development and detection of the coloured stain

In principle, there are two ways to bring the arsine gas into contact with the mercuric bromide-impregnated reagent paper. The older, original approach passed the arsine gas over the surface of a paper strip placed vertically in a narrow tube. Quantification was based on the length of the paper which turned yellow. This method has been widely used, including during the first epidemiological studies of chronic arsenic poisoning in Taiwan. These studies led to the discovery of 'Black foot' disease. Since this approach depends on solid–gas interactions in a rather poorly defined system, it is not surprising that such a set-up cannot give very reproducible results. A critical study by Greschonig and Irgolic [7] showed the weakness of this method. However, this approach has been adopted in the commercially available arsenic field-test kit from Merck (Merck KGaA, Darmstadt, Germany). This system was originally designed for wastewater analysis where arsenic concentrations are much greater than in typical drinking water. The original Merck kit gave an indication of the concentration observed (steps of 0, 100, 500, 1000, 1700 or 3000 $\mu g \, l^{-1}$) and was widely used in Bangladesh. However, early tests by Bangladeshi workers of this and other kits have often been disappointing in the concentration range most needed, i.e. 50–150 $\mu g \, l^{-1}$, e.g. [8]. A more sensitive kit is now available from Merck (Merck 'sensitive') with steps at 10, 25, 50, 100 and 500 $\mu g \, l^{-1}$, but the results of the reliability of this kit are not yet available. Other 'modern' test kits, e.g. the AAN kit (2-19-35 Nagaoka, Minami-ku, Fukuoka, Japan) and the Hach kit, have also brought the minimum detectable concentration down to about 10 $\mu g \, l^{-1}$ and also have greater resolution in the 0–500 $\mu g \, l^{-1}$ range. All of these kits include a calibrated colour scale to enable a semi-quantitative estimate of arsenic concentrations to be made, usually based on comparison with a test strip having a graduated colour scale in six–seven steps: the standard Hach kit has steps at 0, 10, 30, 50, 70 and 300 $\mu g \, l^{-1}$ and a simplified version which has been selected for use in Bangladesh (the 'EZ' Arsenic test kit) has steps at 0, 10, 25, 50, 70, 100, 250 and 500 $\mu g \, l^{-1}$ (50 ml sample) and 0, 35, 75, 175, 500, 1500 and 4000 $\mu g \, l^{-1}$ (9.6 ml sample). Analysis time using these kits is about 30 min per sample and the recurrent cost is less than $1 per sample.

The colour development occurs more efficiently if the gas passes through the reagent paper rather than over it and if the area of paper exposed is small so as to concentrate the available arsine on a small surface area. This approach is adopted by the Arsenator generator.

2.3. Photometry

Most photometric methods can in principle be modified for use in the field. A large number of battery-powered colorimeters and spectrophotometers is already available commercially. However, the two most important laboratory-based photometric methods, the silver diethyldithiocarbamate (SDDC) method [9] and the molybdenum blue method, both have problems when adapted for field use. The SDDC method uses a volatile organic solvent that is difficult to store and transport. The molybdenum blue method suffers from interferences from other anions, particularly phosphate, and while these can be partially eliminated [10], and ultimately they limit the sensitivity of the method. As phosphate is often found at high concentrations in arsenic-affected groundwaters, this is a significant drawback to the method. Furthermore, the large number of reagents involved with both of these methods and the careful cleaning of cuvettes required for precise work make such methods less readily adaptable to field conditions.

2.4. The Gutzeit method

Most available field-test kits for arsenic are presently based on the 'Gutzeit' method. This method was first described in 1879 [11] and used silver nitrate crystals as the detector. Although the silver nitrate crystals were later replaced by a more practical paper strip impregnated with mercuric chloride [12] and [13], and subsequently with mercuric bromide, the method is still referred to by the name of its original inventor, Gutzeit. The Gutzeit method has dominated arsenic testing for more than 100 years. Unfortunately, analytical chemists have not made a great enough effort to replace the toxic mercuric bromide reagent with a less harmful reagent.

The arsine gas replaces bromide attached to the mercuric bromide with an AsH_2 group in a stepwise manner. Mercuric bromide is colourless but the reaction product gives a yellow colour when one arsine group is bound and then becomes progressively more brown as more bromide is replaced. Unfortunately, the human eye is not particularly sensitive to the yellow colour, especially under field conditions. The practical detection limit therefore depends strongly on the ability of the operator to detect the faint yellow stain on the paper. At the beginning of a field survey, the technicians must be tested on their ability to distinguish different intensities of this characteristic colour. Also once the reaction product is formed the yellow colour gradually fades, especially in sunlight. Therefore, comparison with the colour chart must be done as rapidly as possible.

The quantification becomes more reliable when the colour detection is replaced by an instrument. This was first demonstrated in 1944 [14] and can be done in two ways: firstly, by measuring the intensity of the coloured spot by passing light through the reagent paper as in classical photometry, or secondly, by measuring the intensity of the spot using reflected light. Unlike the human eye, such an instrument can use a light source with a narrow range of wavelengths and can therefore be made to be especially sensitive to the yellow colour of the spot. This automation lowers both the detection limits, improves the reproducibility and enables the concentration to be determined on a continuous scale. With such instruments, detection limits ($3s$) of 0.5 µg l^{-1} for the light-through approach and 1–2 µg l^{-1} for the reflected-light approach are possible. Modern electronics makes it possible for such instruments to be the size of a pocket calculator. The original Arsenator used the light-through approach and PeCo75, a miniaturized, hand-held version of the Arsenator, uses the reflected-light approach. Such an instrument can typically measure arsenic concentrations in water ranging from 1 to 300 µg l^{-1}.

For most natural groundwaters, the Gutzeit method is selective enough. Although other elements such as antimony and selenium also form hydrides, it is generally assumed that most groundwater samples do not contain these interfering elements at sufficiently high concentrations to interfere with the determination of the arsenic. In practice, hydrogen sulfide is usually the most serious interferent with this method. This can occur as a dissolved gas in groundwaters that are strongly reducing. A second source of hydrogen sulfide is the dissolved organic matter in the water which can contain sulfur compounds which produce hydrogen sulfide during HG. Hydrogen sulfide is not produced by sulfate reduction in the analytical procedure since the zinc–acid and borohydride systems employed do not reduce sulfate. Hydrogen sulfide gas reacts with mercury to give black mercury sulfide which both makes the mercuric sulfide insensitive to arsenic and produces a grey colour which obscures the arsenic coloration. Even traces of hydrogen sulfide gas lead to a grey spot on the reagent paper. Therefore, if present, it is necessary to remove this interference. This can be achieved by passing the gas through a piece of cotton–glass wool or filter paper impregnated with lead acetate before the gas reaches the mercuric bromide reagent. Alternatively, hydrogen sulfide

can be oxidized to sulfate before HG is begun. This latter approach is rather more time-consuming and needs additional reagents but is adopted in the standard Hach approach where Oxone® (potassium peroxymonopersulfate, $KHSO_5$) is used to oxidize sulfide to sulfate and then EDTA is used to remove the excess Oxone® [5]. This procedure avoids the use of hazardous lead salts and is not affected by up to 5 mg l^{-1} H_2S. However, the Hach 'EZ' kit does not use this procedure and resorts to Pb acetate-soaked cotton balls. From our experience in Bangladesh, hydrogen sulfide is detected by the darkening of lead acetate-soaked glass wool in 10% or fewer groundwaters, and even less frequently by smell. It is normally only present at the micrograms per litre level. Nevertheless, its occasional presence at significant concentrations means that it should be routinely removed to ensure reliable As analyses.

The field-test kits can in principle also be used to determine the arsenic speciation, specifically the concentration of As(III) and As(V) species, albeit only at a semi-quantitative level. This depends on the different kinetics of hydride formation between As(III) and As(V) species—as noted above, As(III) forms the hydride more rapidly than As(V) using the traditional $KI+SnCl_2+Zn+HCl$ reduction method. The kits do not normally detect organic As species, which in any case are not normally significant in groundwaters, but can be made to do so if a preliminary oxidation step is included.

2.5. Quality assurance and the management of large screening programmes

An important aspect of analysis is quality assurance which is necessary not just to confirm blank values but also to test the accuracy of the method periodically and to provide feedback to the survey manager and testers on the quality of the results. The easiest way to do this is to check the results of a standard solution. Deionized or naturally low-As water can be spiked with different concentrations of arsenic and the recovery is used to indicate how well the analysis is performing. An alternative approach is to analyse a proportion of samples in duplicate. Depending on the extent to which such quality control is considered important and the facilities available, a certain number of samples can also be usefully sent to a laboratory for cross-checking by instrumental analysis. In this case, it is important that the samples sent to the laboratory are well preserved by acidification. Such checking is especially important at the beginning of field surveys in order to detect errors in kits and procedures at an early stage. It is a good practice to disguise and randomise all check samples before submission for QA testing.

Since large field surveys normally involve a large number of people, it is important that there are clear lines of authority and assigned responsibilities. It is also necessary to have reliable sample tracking procedures in place. These management procedures should be documented in the sample protocols. Care should also be taken to label clearly and indelibly all samples sent to the laboratory.

3. Sampling and analytical strategy for mass screening

3.1. Sampling

Sampling procedures are always a critical part of environmental assessment. The result of a chemical analysis is no better than the sample on which it is based. Field-testing has several major advantages: there is no need for transport, no storage and therefore no need for

preservation, e.g. acidification to minimize adsorption on the surface of the bottle and prevent precipitation of hydrous iron oxides which can adsorb As. Also, with field-testing, the well owner can be informed of the result relatively easily and rapidly. In Bangladesh and other hot countries, attempting to keep samples cool over a long period during transport to a laboratory can be difficult. On the other hand, laboratory testing has obvious advantages in terms of instrumentation and QA.

There is some debate amongst geochemists and water quality planners over whether the water sample should be filtered or not. It is a normal practice in most geochemical studies of groundwater to filter the sample in order that only 'dissolved' substances are measured. A 0.2 or 0.45 μm membrane filter is usually used for this. This filtering is done so that speciation calculations, such as mineral saturation indices, can be carried out. Even so, it is acknowledged that there are sometimes particles smaller than 0.2 μm and that these will not be removed with such filters. Therefore, care has to be taken whenever analytical results are used in such calculations, filtered or not. Acidification of samples for preservation purposes is likely to at least partially dissolve mineral particles if present, and this will lead to an overestimate in the dissolved concentrations reported. Acidification of unfiltered samples is therefore not usually undertaken or recommended.

For drinking-water compliance, the presence of particles and their adsorbed and coprecipitated load can make a difference to the total ingested load and must therefore be considered. Again, the speciation is important: if the trace metals can be desorbed or dissolved in the gut, then they should be included in the measurement. If they pass straight through the gut unchanged, then it makes more sense not to include them. In practice, many particles could partially dissolve in the gut. WHO [15] are not prescriptive on the need to filter. In practice, most water samples for compliance testing are not filtered, i.e. the raw water is used, and this is recommended for Bangladesh mass screening surveys. Clearly, when the samples contain high concentrations of particles, this may upset the instrumentation (by clogging), alter the apparent concentration (e.g. due to scattering in colorimetric methods) and lead to highly variable results. It is therefore important to record the presence or absence of turbidity, and preferably, the analysis of visibly turbid samples should be avoided. Extreme turbidity can be a problem with wells that have not been pumped for some time, but the turbidity usually clears with use and so turbidity is not usually a problem with freshly collected groundwater samples from actively used wells. We discuss an example of the impact of filtering on Bangladesh groundwaters below.

For geochemical purposes, it is usually desirable to analyse water representing the water present in the aquifer, and so the extent of purging of the borehole and pipework prior to sampling is an important part of the sampling protocol. It is often specified in sampling protocols that at least three water column volumes should be purged before sampling [16]. This is easy to do with wells fitted with an electric or diesel pump but less easy with hand-pump tubewells, especially when the wells are deep (the depths of 'deep' tubewells in Bangladesh can exceed 200 m). For the typical situation in Bangladesh, a simple rule of thumb is to apply one stroke of the hand-pump for every foot of well depth. Clearly, the significance of this purging will depend on the extent to which the well has been used before sampling.

It is also important to know the construction of the borehole, specifically the location and length of the well screen. In shallow tubewells in Bangladesh, there tends to be one length (6 ft or 2 m) of plastic-well screen at the base of the well and in deep wells, twice this length of

screen. Deep tubewells in Bangladesh are sometimes screened at several depths in order to maximize the usable yield. This can greatly affect the observed arsenic concentration in the groundwater and must be considered when interpreting the analytical results in terms of geochemical processes, i.e. 'deep' wells cannot be assumed to derive all of their water from the 'deep' aquifer.

Because the time required for the measurement of arsenic is usually less than that required for sampling, it is more appropriate to establish a good working place to make the analytical measurement (i.e. a temporary 'field lab') and to send out samplers to collect samples. The density of tubewells in much of Bangladesh is so high that the distances involved are short and that the delay due to transport of the sample to the 'field lab' should not seriously affect the arsenic analysis. Any ferrous iron that has precipitated as an iron(III) oxide floc on air oxidation is readily redissolved under the acidic conditions under which the arsenic analysis normally takes place. As we have discussed earlier, the methods used for arsenic determination generally result in a light-sensitive compound. Therefore, the work place should also be out of direct sunlight.

Where there is interest in the geochemical source of the arsenic, it is also worthwhile collecting ancillary field data, e.g. electrical conductivity, alkalinity, pH, redox potential and dissolved oxygen as well as additional samples for more comprehensive chemical analysis. It is also important to record details of the sample, such as smell, colour, presence of suspended matter or/and turbidity and bubbles, and of the well type, owner and location.

3.2. Field-testing versus laboratory testing

There has been a considerable expansion in the analytical facilities for arsenic testing in Bangladesh, particularly by the private sector, since 1997. While widespread laboratory testing is advocated by some, there are three major obstacles at present that make this difficult to achieve: firstly, the lack of laboratories with the analytical capability for sufficient sample throughput and a track record in quality control; secondly, the lack of management experience to organize the sampling, sample tracking and reporting of results on the scale required; and thirdly, logistical problems associated with the transport of samples from the field to the laboratory. This makes field-testing more attractive than laboratory testing, at least in the short term.

Furthermore, a recent interlaboratory comparison of arsenic analysis of eight synthetic standards and two groundwater samples by 17 Bangladeshi laboratories has proven disappointing. Less than one third of the participating laboratories were within 20% of the expected values [17]. Reported concentrations for the two groundwater samples having arsenic concentrations of 61 and 68 μg l^{-1} as measured by ICP–MS at the University of Rochester, USA, ranged from 0 to 396 μg l^{-1}. The participating laboratories used their normal analytical procedures. Not surprisingly, some laboratories were consistently better than others. Therefore, even laboratory testing does not guarantee reliable results. It is clear that there is a need for some kind of continuous round-robin quality control programme in Bangladesh to help the laboratories raise their standards of analysis and to give funding agencies confidence in the quality of results obtained. This exercise could be accompanied by a formal accreditation scheme, or could be undertaken more informally. Such programmes exist in many western countries, and are run on a commercial basis.

Ideally a mass screening programme based on laboratory analysis requires automatic samplers linked to sophisticated instruments such as HG–AAS capable of carrying out several hundred analyses per day. Planning and implementing such a programme raise many new and important issues, including those of responsibility and leadership, and require that many decisions need to be made, both technical and non-technical. The analytical approach to adopt is one such decision. Transporting large numbers of samples to a distant laboratory, analysing them reliably at the trace (μg l^{-1}) level and in a timely manner, reporting back the results to the well owner and those responsible for recommending alternative water supplies all provide many opportunities for introducing error, confusion and delay. Field analysis avoids many of these problems but will of course create new ones, particularly how to ensure a consistent quality of results. Most screening programmes using field-test kits have used laboratory analysis to check a proportion, typically 2–10%, of the field tests. This seems a reasonable approach, providing that there is rapid feedback from the quality control laboratory to the field workers so that lessons can be learnt and any corrective action required is taken. Needless to say, quality control in the chosen QC laboratory itself has to be carefully maintained again by regular cross-checking with recognised standards.

Some desirable features of the analytical approaches apply to both field and laboratory methods. The most important are reliability and robustness, accuracy, including bias and precision, sensitivity and selectivity, cost-effectiveness, the time needed for a single determination, personnel safety and potential environmental impact. Reliability and robustness include not only the ability of the equipment to withstand the rigours of continuous field use but also whether the method is simple enough to be operated reliably and consistently by often relatively unskilled technicians (especially in the field). Since field kits have to be used in very large numbers, there is relatively little time to train and supervise field workers. Staff turnover of testers can also be high. The hot, humid climate found in tropical countries also has to be taken into account in terms of the reliability of equipment and chemicals, i.e. the equipment must be 'ruggedised' and the chemicals should be available in individually sealed and dated packets.

The usefulness of a given field-test kit is determined in part by the amount of information it provides. At the lowest level, such kits only indicate the presence of arsenic above a certain threshold concentration. This was characteristic of many of the earlier 'yes/no' kits used in Bangladesh. A coloured stain was either recorded as visible or not visible. The detectable arsenic concentration of these kits was usually conveniently said to be 50 μg l^{-1} (the Bangladesh standard), but blind testing has often shown that the cut-off was considerably higher, sometimes as great as 100 μg l^{-1}. The original Merck field-test kit, widely used in Bangladesh, had a stated minimum detectable concentration of 100 μg l^{-1}. While such sensitivity was adequate for early surveys aimed at assessing the regional pattern of arsenic contamination and the scale of the problem, as for example the DPHE/UNICEF survey carried out in Bangladesh in July 1997, it is not acceptable for compliance testing where important decisions have to be made about the safety of a particular water source and where the standards in place are invariably lower than 100 μg l^{-1}. Such compliance testing demands a higher degree of sensitivity and accuracy, particularly where a well has been installed at considerable expense and has obvious benefits in terms of convenience of use and microbiological safety. To condemn forever such a well for drinking water use (but not for washing etc.) is a serious step to take. Fortunately, the latest field-test kits described above are both more sensitive and have a greater resolution than earlier versions.

Field use of instruments such as the Arsenator offer a 'half-way house' between field and laboratory instruments. In principle, such approaches could have the accuracy and precision generally given by sophisticated laboratory methods but in the field. Fig. 2 shows a comparison of the field and laboratory results from 211 samples collected in Mandari village, Lakshmipur, south-eastern Bangladesh during November 1999. Arsenator 510 was used for the field measurements. These measurements were made on unfiltered and unacidified samples. The laboratory results were made on filtered (Whatman 0.2 μm membrane filter) and acidified (1% nitric acid) samples in the BGS (Wallingford) laboratories using direct aspiration into a Perkin Elmer ICP–AES (no HG). While ICP–AES used in this way has a rather high detection limit ($3s_0$) for arsenic (about 20 μg l^{-1}), the majority of samples exceeded this concentration, and the results show clearly that the field method can give similar results to those found in the laboratory. Also since the laboratory sample was based on a filtered sample (0.2 μm) and the field sample was unfiltered, the lack of a systematic difference between the two sets of results indicates that the filtering did not significantly affect the results. Therefore, it appears that in the tested wells, particles greater than 0.2 μm diameter are not a common source of arsenic in these well waters. Smedley et al. [18] also found that there was little difference between filtered and unfiltered water in 29 samples from Chapai Nawabganj, Bangladesh (most agreed within ±20%).
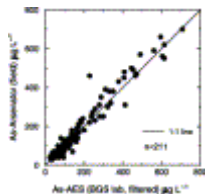

Fig. 2. Comparison of arsenic analysis of groundwater samples from Mandari, SE Bangladesh based on field-testing with an Arsenator 510 (unfiltered sample) and laboratory analysis by direct aspiration ICP–AES (filtered sample).

3.3. Compliance testing and the consequences of analytical error

There are plans to analyse each of the 6–11 million tubewells in Bangladesh at least once. Currently some 10% of tubewells are believed to have already been analysed, and the plans are to analyse the remainder as rapidly as possible. The only feasible approach is through the use of field-test kits, and given the large number of wells involved and the urgency with which the contaminated wells need to be located, initially only one test will be carried out per well. Therefore, the impact of analytical errors are of considerable concern, and the organisations involved (Government of Bangladesh, NGOs, donor agencies, World Bank and UNICEF) have attempted to minimise these errors by evaluating the various kits beforehand, selecting the most reliable and instigating some kind of ongoing QA protocol.

Whatever the characteristics of a test kit, it is necessary that they have an acceptably low probability of giving false-positive and -negative results. A false-positive result records a sample as exceeding the drinking water standard when its true concentration is less than the standard; a false-negative result records a sample as below the standard when it exceeds the standard. While a false-positive result is less serious from a health point of view, it means that the well owner is expected to take some action in order to comply with the standard either by abandoning the well for drinking and cooking purposes or by treating the water. In highly contaminated areas, a false-positive test undervalues what could be a valuable local resource. For example, it would be wasteful to sink a deep tubewell (which in Bangladesh

tends to be acceptably low in arsenic) at, perhaps, 10–20 times the cost of a shallow tubewell if there is a shallow well nearby with acceptable quality, and which in principle could be shared by the community. A false-negative result leads the well users to believe that the water is safe when it is not. Both types of misclassification have a personal cost and detract from the benefits that the well owner expected from their original investment.

It is therefore important that field kits are as accurate as possible. Perhaps more importantly, the kits should behave consistently and reliably under field use so that the inherent errors in the kits are known from the outset and their likely impact can be assessed. It is also obviously important that the kits are used in the way intended, and so as stressed earlier, monitoring of the results is also important.

The range of arsenic concentrations in Bangladesh groundwaters is very large, ranging over more than four orders of magnitude [2]. A relatively large error may therefore be acceptable for some purposes. Certainly, a priority should be to locate all of the very high arsenic wells as rapidly and reliably as possible—say those wells with a concentration exceeding 200 μg l^{-1}.

Knowing the reliability of the instrumentation used, it is possible to simulate the effect of analytical errors on the outcome of a screening survey. For example, 27.43% of all shallow (less than 150 m deep) tubewells sampled in the BGS/DPHE survey of Bangladesh tubewell waters exceeded the Bangladesh standard for arsenic in drinking water (50 μg l^{-1}). This statistic is based on laboratory analysis by BGS, largely carried out by HG–AFS. If we assume that this survey accurately reflects the true distribution of arsenic concentrations in Bangladesh, then we can use a Monte Carlo approach to calculate how the number of non-compliant wells changes as we impose different random analytical errors on the analysis of each sampled well. Following Thompson and Howarth [19], we assume that the analytical precision follows the following model:

(1)
$$s_c = s_0 + k \times c_r$$

where s_c is the standard deviation of replicate analyses having a total arsenic concentration, c. s_0 and k define how s_c varies with concentration. s_0 represents the standard deviation at zero concentration, while k defines a constant relative error. The instrumental detection limit is often defined as $3s_0$. Most analytical methods are dominated by the relative error or k term, but at low concentrations, the s_0 term becomes important. This is especially true where there is a cut-off in detection at low concentrations, for example with field-test kits where the human eye cannot reliably detect the very faint yellow colour produced at low concentrations. Approximately 95% of tests should fall within $\pm 2s_c$ of their true values.

The values of the s_0 and k parameters given in Eq. (1) can be established from the variation in replicate analyses over a range of concentrations. This can either be carried out in a preliminary systematic study of analytical precision at various concentrations or by monitoring the absolute difference between duplicate samples as part of a routine QA programme during the survey [19]. There is little published data to establish these parameters for most laboratories or most field-test kits. Therefore, we have used USEPA Water Supply (WS) Study Performance Data to establish possible analytical errors based on laboratory analyses [20]. These QA data were derived from a large number of replicate arsenic analyses based on a variety of laboratory procedures (AAS with furnace, AAS with HG–ICP, ICP–MS and spectrophotometry using SDDC) (Fig. 3). After the removal of one outlier from this

dataset, linear regression gave s_0=0.30 µg l^{-1} and k=0.088. This corresponds to a detection limit ($3s_0$) of about 1 µg l^{-1} which is reasonable. The only suitable data that we could find for field-test kits were those for the AAN kit in which 100 and 200 µg l^{-1} As(III) standard solutions were tested by 77 Miyazaki University students with no prior experience of using the test kit [21]. This gave a standard deviation of 11 µg l^{-1} at 100 µg l^{-1} and 61 µg l^{-1} at 200 µg l^{-1}. We therefore estimate that s_0=5 µg l^{-1} and k=0.2 ('AAN kit'). This corresponds to a detection limit ($3s_0$) of 15 µg l^{-1}. We have no detailed precision data for the Hach kit, but Kroll (Dan Kroll, Hach Company, personal communication, 2002), using a Hach kit and a single operator using three representative lots of reagents and 50 tests of a 10 µg l^{-1} standard solution, found no results that were not properly matched to the 10 µg l^{-1} colour dot within $\frac{1}{2}$ step, i.e. all results were within the range 5–20 µg l^{-1}. We can therefore assume as a worst case that s=10 µg l^{-1} at 10 µg l^{-1}. In the absence of any precision data for higher concentrations, we assume that k=0.3 which gives s_0=7 µg l^{-1} ('Hach kit'). Precision testing of PeCo75 has been reported by BAMWSP/DFID/WaterAid Bangladesh [22]. Repeat testing of 50 µg l^{-1} (n=57) and 300 µg l^{-1} (n=54) standards gave standard deviations of 8.4 and 43.7 µg l^{-1}, respectively, giving k=0.14 and s_0=1.4 µg l^{-1}. This gives an error between that from sophisticated laboratory procedures and that from field-test kits. These results are summarised in Table 1.
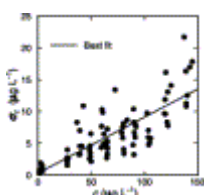

Fig. 3. Variation of the standard deviation of replicate arsenic measurements versus arsenic concentration for a variety of laboratory-based methods based on data returned to the USEPA by various laboratories participating in water supply analyses.

Table 1. Various estimates of the parameters for the Thompson–Howarth error model, Eq. (1), for arsenic analyses based on various analytical methods

Analytical method	s_0 (µg l^{-1})	k	Source of data	Reference
Various laboratory procedures	0.30	0.088	USEPA water supply data from various laboratories	[20]
PeCo75 (a hand-held 'Arsenator')	1.4	0.14	Based on duplicate analysis of 50 and 250 µg l^{-1} As standards by three survey teams every 2 days during a survey	[22]
AAN field-test kit	5	0.2	Replicate analysis of 100 and 200 µg l^{-1} As standards by 77 university students	[21]
Standard Hach field-test kit	7	0.3	'Pessimistic' parameter estimates based on 50 analyses of a 10 µg l^{-1} As standard	See text

In practice, there may be other sources of analytical error including operator errors arising from the lack of ideal conditions in the field and analytical bias, and so these error estimates may be if anything too low. The magnitude of sampling errors arising from temporal

variations of the well-head chemistry is largely unknown and also contributes to the overall sampling error. Here we only consider random error and strongly recommend that a quality assurance programme be undertaken to monitor and control the actual accuracy achieved during a survey.

As demonstrated above, the probability of misclassification not only depends on the analytical errors but also on the true As concentration. In a simple two-way classification (compliant, non-compliant), the possibility of misclassification is greatest at the cut-off value, since there is a 50% probability of misclassification at this concentration irrespective of the analytical error (but assuming some error). The standardized analytical error is given by

(2)
$$z = \frac{t-c}{s_c},$$

where t is the cut-off value and c is the true As concentration. The probability of misclassification, $P(c)$, is then given by

(3a)
$$P(c;\ c \leq t) = 1 - \Phi^{-1}(z|c \leq t)$$

(3b)
$$P(c;\ c \geq t) = \Phi^{-1}(z|c \geq t),$$

where $\Phi^{-1}(z)$ is the inverse standard normal distribution function.

$P(c)$ can be plotted against c for both laboratory- and field-based procedures, given the error models described above (Fig. 4). With the more precise laboratory testing, few samples (<5%) with concentrations outside the 40–60 µg l^{-1} range will be misclassified, but with the less precise field-test kit, 50% of wells in this range are likely to be misclassified. With the 'AAN' field-test kit, wells containing 60 µg l^{-1} As have a 28% probability of testing as acceptable (i.e. below 50 µg l^{-1}), whereas wells containing 100 µg l^{-1} have a 2.2% probability of testing as acceptable.

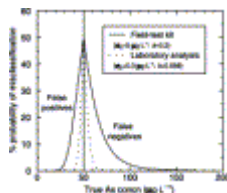Fig. 4. Percent probability that a tubewell having a given arsenic concentration will be misclassified according to the 50 µg l^{-1} arsenic standard assuming different error models for the arsenic analysis ('field-test kit' and 'laboratory analysis') (based on (2), (3a) and (3b)).

The expected error was simulated for each of 3208 shallow groundwater samples 1000 times, assuming the error to be randomly derived from a Gaussian distribution with s_c calculated according to the above error model. The overall impact of these errors on the extent of well compliance was then calculated (Table 2). Three levels of compliance were considered: 10 µg l^{-1} (the WHO guideline value), 50 µg l^{-1} (the Bangladesh drinking water standard) and 200 µg l^{-1} (a seriously high concentration).

Table 2. Estimated effect of analytical error on the misclassification of shallow tubewells in Bangladesh (61 districts)

Error model (µg l^{-1})	Laboratory analysis (s_c=0.30+0.088[As$_T$])	Field-test kit (s_c=5+0.2[As$_T$])
True % of wells exceeding 10 µg l^{-1}=45.95% (*compliance criterion*=10 µg l^{-1})		
% wells as false-positive	0.45	4.22
% wells as false-negative	0.46	2.79
% wells misclassified	0.91	7.01
% wells exceeding criterion	45.93	47.38
True % of wells exceeding 50 µg l^{-1}=27.43% (*compliance criterion*=50 µg l^{-1})		
% wells as false-positive	0.42	1.37
% wells as false-negative	0.58	1.86
% wells misclassified	0.99	3.23
% wells exceeding criterion	27.27	26.93
True % of wells exceeding 200 µg l^{-1}=9.88% (*compliance criterion*=200 µg l^{-1})		
% wells as false-positive	0.39	0.91
% wells as false-negative	0.39	1.06
% wells misclassified	0.78	1.97
% wells exceeding criterion	9.89	9.73

With the error model based on laboratory analyses, the calculations suggest that approximately 0.78–0.99% of wells will be misclassified at each of the three compliance levels with approximately equal numbers of false-positive and false-negative results. This means that the estimated overall percentage of non-compliant wells is very close to the true number, even though about 1% have been misclassified. While 1% appears to be a small value, with perhaps 11 million wells in Bangladesh, it still suggests that some 100 000 wells could be misclassified. The chance of misclassification is considerably greater with the field-test kits. Given the assumed precision of the 'AAN' field-test kit, some 2–7% of wells are estimated to be misclassified, with the highest risk of misclassification being when the cut-off concentration was set at 10 µg l^{-1} rather than at 50 or 200 µg l^{-1}. This increases to 12% for the Hach kit (detailed results not shown). There are somewhat more false-positive results with a 10 µg l^{-1} compliance level than false-negative results, while at a compliance level of 50 µg l^{-1}, the reverse is true. The number of wells misclassified at the 200 µg l^{-1} level with the field-test kits is relatively small but so are the number of wells exceeding this value. Some 11–16% of these high-As wells would be classified as less than 200 µg l^{-1}, whereas in fact they exceed 200 µg l^{-1}. However, very few (<0.6%) of these wells would be reported as less than 50 µg l^{-1}. Therefore, given the assumed precision of the field-test kits, they appear to be reliable in being able to identify the most contaminated tubewells. Experience confirms this. The groundwater arsenic map of Bangladesh based on the 1997 DPHE/UNICEF field-kit

survey shows a similar location for the most contaminated areas compared with the laboratory-based analyses of the 1998/1999 DPHE/BGS survey. However, the field-kit survey does not resolve the distribution of As in much of northern Bangladesh where As concentrations are frequently below 50 µg l^{-1}. There has been a considerable improvement in the field-test kits since 1997 and modern kits with the precision of the 'Hach', 'Merck' or 'AAN' kits would now give a better spatial resolution in northern Bangladesh.

The reliability of the field-test kits depends not only on the compliance level chosen but also on the distribution of concentrations found in the survey area. For example, at the 50 µg l^{-1} level and based on the assumed precision for the standard 'Hach kit', the extent of misclassification in the 12 worst-affected districts (5.5%) is slightly greater than in the country as a whole (4.8%), while at the 10 µg l^{-1} level, the reverse is true: whole country (11.6%) and worst-affected districts (5.1%). This reflects the proximity of the compliance concentration to the median concentration (135 µg l^{-1} in the 12 worst-affected districts compared with 6 µg l^{-1} in the country as a whole).

Using the shallow groundwater data, the number of misclassified wells can be seen to be directly related to the value of s_0 and the value of the relative error parameter, k (Fig. 5). Therefore, improvements in field-test kits can be measured by the reduction in the s_0 and k parameters and can be evaluated in terms of the consequent reduction in the percentage of misclassified wells. In these calculations, slightly more than half of the misclassified tests are false-negatives. On a national scale, it appears that with the present error model for field-test kits, less than 5% of the wells will be misclassified. In districts where a greater percentage of tubewells is close to the 50 µg l^{-1} standard, a somewhat greater percentage of wells will be misclassified following Fig. 4.

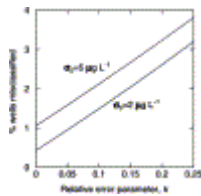

Fig. 5. Plot of the percentage of Bangladesh shallow tubewells that will be misclassified according to a 50 µg l^{-1} arsenic standard as a function of the measurement error as reflected by the values of the parameters in the assumed error model (Eq. (1)).

In one sense, such calculations are reassuring in that they show that even with analytical methods having quite large errors, as with the field-test kits, the majority of wells exceeding the Bangladesh arsenic standard should be identified correctly. On the other hand, the absolute number of false-positive and false-negative results is still high, bearing in mind the large number of wells that must be analysed, and in areas with a large number of wells close to or exceeding the compliance limit, a relatively high proportion of wells will be misclassified. It is clear that every effort should be made to ensure as reliable analyses as possible. However, it is equally important to appreciate that there will inevitably be analytical errors whatever method of analysis is used and that the consequences of these errors should be appreciated from the outset. Also since time is important, it is not possible to delay testing until the 'ideal' field-test kit is available.

These results raise the question as to how such uncertainty should be communicated to the well owner. The normal response in Bangladesh is to paint a well red (non-compliant) or green (compliant) depending on the test result. This will inevitably result in a considerable number of wells being misclassified because of the errors in the test as outlined above. Our simulations suggest that the number of false-negative results is likely to exceed the number of false-positive results. The extent of misclassification will become apparent as repeat testing is carried out. As the better field-test kits already provide a semi-quantitative estimate of arsenic concentration, it has to be decided whether this additional information should be given to the well owner or not. In general, it seems best to provide the well owner with as much information as possible along with advice on its safety for drinking. While all non-compliant wells should be dealt with, the priority must be to identify and deal with the most contaminated wells first.

Analysing each well in duplicate would reduce the extent of misclassifcation appreciably but has a high penalty in terms of time and cost. There are also persistent rumours in Bangladesh that tubewell arsenic concentrations vary with time. These are as yet largely unsubstantiated, and there is usually no indication of whether these variations are random, or reflect a long-term trend, or both. The indication usually given is that concentrations are increasing with time, often quite dramatically. While this seems unlikely as a generalization, it has led to the call for 'annual' monitoring of each tubewell. Aside from the large resources required to do this, it is not clear how the variation will be resolved into a possible time-varying component and measurement error, at least on an individual well basis. The scope for confusion is considerable. Resolving the issues of time dependence requires systematic monitoring with the highest possible precision measurements, i.e. laboratory analyses.

While all the evidences point to a large degree of short-range (over 1 km or so) variability in tubewell arsenic concentrations in Bangladesh, there is likely to be some degree of short-range spatial dependence. This may help to highlight locally anomalous wells that would then merit priority for retesting. Furthermore, if well owners are given as much information about their test result as possible then they can judge how to act. For example, there is an important difference between a well that just fails, say at 60 μg l^{-1}, and one that fails by a far wider margin, say at 300 μg l^{-1}.

4. Conclusions

-

  The large number of tubewells that need to be tested for arsenic in Bangladesh (perhaps as many as 11 million) and elsewhere means that field-test kits offer the only plausible approach for mass screening. It is also only feasible to carry out one test per well, at least initially.

-

  Most existing field-test kits for arsenic are based on the classical Gutzeit method which generates arsine gas by reduction and detects it by the strength of the yellow-brown colour produced on mercuric bromide-impregnated paper. There have been significant recent improvements in the kits since 1997, notably by improving the

contact between arsine gas and impregnated paper. This has enabled the detection of As down to approximately 10 µg l^{-1}.

- Other recent improvements include the replacement of liquid acid with a solid acid (sulfamic acid), the replacement of zinc by sodium borohydride and the development of an electronic device to measure the yellow-brown colour of the stain in a quantitative manner.

- The challenge is to produce field-test kits which are robust, reliable, cheap and simple enough to be used by relatively unskilled technicians in Bangladesh and elsewhere. These, and their supplies, should be readily available in the local markets.

- The precision of field or laboratory testing equipment should be established over a range of As concentrations by measuring the precision of a range of standards (say 0, 50 and 300 µg l^{-1}) and by duplicate analysis during screening. The results should be fitted to the Thompson–Howarth error model. The aim should be to ensure that the quality of analyses found during an actual screening programme is kept as close as possible to that achievable under ideal conditions, i.e. to achieve the full design potential of the equipment.

- By establishing a realistic 'error model' for the analytical method to be used in a screening survey (either field- or laboratory-based) and having some prior information about the likely distribution of arsenic concentrations expected (from a random reconnaissance survey), it is possible for planners to estimate the likely extent of misclassifications during the screening survey.

- Since there are known to be a large number of wells in Bangladesh with arsenic concentrations close to the existing national drinking water standard of 50 µg l^{-1}, a significant number of misclassifications (false-positives and negatives) are inevitable and should be expected even when laboratory analyses are used. However, even with the field-test kits now available, these errors could be brought down to no more than 5% of the wells tested on a national scale.

- Giving private well owners as much information as possible about the test result will alert them to the likely degree of contamination of their water supply and enable them to make their own decisions accordingly.

Acknowledgements