The key concerns when changing site sampling methods are (1) whether the results acquired using the new method will be substantially the same as those acquired by the previously used and accepted method and (2) whether the regulators will accept results acquired by the passive sampling method. Various media can be sampled via passive sampling. Groundwater sampling is subject to the most constraints when evaluating and comparing the data collected. However, many of the considerations and methods described in this section could be applied across all media.

4.1 Site Data Quality Objectives

Before undertaking an evaluation of the results between sampling methods, the site DQOs should be reviewed to determine how the sampling results are used in site decision-making, the key points of comparison between the existing and new method, and what the regulators want to see to allow a change in sampling method. In most cases it is a simple process to discuss the evaluation objectives with the regulators up front so that criteria can be developed prior to beginning an evaluation.

4.1.1 Project-Specific Criteria

Methods used to compare the data should be based on project objectives. For example:

- If groundwater sample data are being used to determine whether, or to what extent, a site has specific chemicals, the comparison may be focused on whether both active and passive techniques indicate similar concentrations at low levels across a wide range of chemicals.
- If the data are part of a long-term monitoring program, the comparison may be specific to whether the different sampling methods lead to the same decision, based on exceedance of regulatory screening levels or criteria for a known set of chemicals.
- A comparison of monitoring data at an active remediation site may be more directed toward the general changes and trends in the concentration of a limited number of chemicals within a treatment area, rather than having agreement on achieving chemical action levels.

4.1.2 Field Data Collection Requirements

Field data collected on site can be used to compare and support the method transition. Sampling results should be evaluated in the context of other field factors that can influence the sample results. A project-specific plan should consider site-specific field data, hydrogeologic differences, and additional information that will help inform whether data variability may be attributable to factors other than the change in method. Following QA/QC procedures may help account for some of these factors, including:

- Physical factors: groundwater elevation, well/probe construction details, tidal influences, seasonality, sampling depth, weather conditions
- Geochemical factors: medium temperature, pH, turbidity, oxidation-reduction potential (ORP), aerobic/anerobic conditions, dissolved gases
- Other factors: vandalism, user experience, equipment malfunction, equipment fouling

4.2 Results Comparison Methods

Below are three techniques for comparing results that can be effective when considering changing sampling methods.

- Historical comparison: Sample using the proposed (passive) technique and compare the results to historical data. This is the least costly method of comparison and may be suitable when there are long-term, consistent, and stable data available.
- Bracketed comparison: Sample some of the locations by alternating between the proposed (passive) and current (active) sampling methods for three or more rounds of sampling. This strategy provides results from the passive method that are “bracketed” between two active sampling results occurring before and after the passive result. Although samples are not taken contemporaneously, changes in detected chemicals or concentration trends may be noted and evaluated. This method takes longer but is less costly than side-by-side evaluations.
- Side-by-side comparison: Perform the proposed (passive) and the current (active) sampling methods

sequentially during a single sampling event to ensure equivalent sample conditions. The passive sampler should be deployed in advance of the scheduled sampling event (to account for sufficient minimum residence time). On the sampling date, the passive sampler is recovered, and immediately after, the active method is implemented, and a sample is collected. Due to the collection and analysis of two samples, this comparison method will be more costly. Because of time and cost considerations, side-by-side evaluations are usually employed at a representative set of locations, rather than all the sampling locations.
When conducting side-by-side groundwater comparisons of active sampling to passive sampling methods, similar results would be expected in wells with 5- to 10-foot screens, unless there were exceptional hydrogeologic differences in the borehole. As screens get longer than 10 feet and the hydrogeologic or geochemical conditions vary, results may vary somewhat between active and passive methods. When contaminant concentrations are variable, the differences in results can usually be explained by further study of the local hydrogeologic and geochemical conditions.

4.3 Result Comparisons between Sampling Technologies

What methods will be employed to compare each data pair?

The U.S. Geological Survey (USGS) provides guidance for groundwater on how to evaluate the data from a side-by-side sampling event, suggesting the following general guidelines for acceptable relative percent differences (RPD) between sample concentrations (Imbrigiotta and Harte 2020 [9LSVPE48] Imbrigiotta, Thomas, and Philip Harte. 2020. “Passive Sampling of Groundwater Wells for Determination of Water Chemistry.” In U.S. Geological Survey Techniques and Methods. https://doi.org/10.3133/tm1d8.). RPD is a common statistical tool used to compare two data points in side-by-side sampling evaluations for technology usability for a site.

The USGS recommended the following RPDs based on chemical concentrations:

- RPD up to +/- 25% VOCs & trace metal concentrations > 10 ug/L
- RPD up to +/- 50% for VOC & trace metal concentrations < 10 ug/L
- RPD up to +/-15% major cations & anions concentrations mg/L range

Lower RPDs mean the two data points are similar. RPDs begin to fail as a practical comparison when concentrations are low. For example, comparing 2 µg/L to 5 µg/L yields a difference of 3 µg/L, which for many regulated chemicals would not be a significant difference that leads to different site decisions. In this example, the calculated RPD is an unacceptable 86%. Therefore, in these cases of low concentration results (for example, within several times the quantitation [or reporting] limit), other evaluation techniques may be appropriate, such as comparison of the two method results and absolute difference between them to the target chemical’s project screening value.

The Techniques and Methods 1–D8 USGS publication also states “one of the more effective ways to compare concentration results” is to plot the data on a 1:1 correspondence on an X-Y plot with the passive results on one axis and the active results on the other axis (Imbrigiotta and Harte 2020 [9LSVPE48] Imbrigiotta, Thomas, and Philip Harte. 2020. “Passive Sampling of Groundwater Wells for Determination of Water Chemistry.” In U.S. Geological Survey Techniques and Methods. https://doi.org/10.3133/tm1d8.). “If the two sampling methods collect the same concentrations, the points will plot on or close to the 1:1 correspondence line” (Imbrigiotta and Harte 2020 [9LSVPE48] Imbrigiotta, Thomas, and Philip Harte. 2020. “Passive Sampling of Groundwater Wells for Determination of Water Chemistry.” In U.S. Geological Survey Techniques and Methods. https://doi.org/10.3133/tm1d8.). Outliers may represent well-specific anomalies such as turbidity.

To determine statistical confidence intervals around the data comparison, standard linear regression methods may be applied depending on normality of data sets. Other nonparametric-based methods—for example, Passing-Bablok regression (Passing and Bablok 1983 [YM82DEUZ] Passing, H., and null Bablok. 1983. “A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods. Application of Linear Regression Procedures for Method Comparison Studies in Clinical Chemistry, Part I.” Journal of Clinical Chemistry and Clinical Biochemistry. Zeitschrift Fur Klinische Chemie Und Klinische Biochemie 21 (11): 709–20. https://doi.org/10.1515/cclm.1983.21.11.709.; Passing and Bablok 1984 [WTFATPGG] Passing, H., and W. Bablok. 1984. “Comparison of Several Regression Procedures for Method Comparison Studies and

Determination of Sample Sizes. Application of Linear Regression Procedures for Method Comparison Studies in Clinical Chemistry, Part II." Journal of Clinical Chemistry and Clinical Biochemistry. Zeitschrift Fur Klinische Chemie Und Klinische Biochemie 22 (6): 431–45. https://doi.org/10.1515/cclm.1984.22.6.431.) or Lin's concordance correlation coefficient (Lin 1989 [JLGS7F6P] Lin, L. I. 1989. "A Concordance Correlation Coefficient to Evaluate Reproducibility." Biometrics 45 (1): 255–68.; McBride 2005 [C7KBPCYJ] McBride, GB. 2005. "A Proposal for Strength-of-Agreement Criteria for Lin's Concordance Correlation Coefficient." NIWA Client Rep. https://www.scienceopen.com/document?vid=46b1768a-002f-48fa-a238-673d8373ee6a.)—may also be used to understand comparability and usability of results. The appropriateness of specific statistical methods and acceptance criteria for data usability are expected to vary by constituent, the sampling methods compared, and project DQOs.

4.4 Other Comparison Considerations

There are a few things that should be considered when comparing the results from sampling events.

- Do the data appear to follow the trend from the past several active sampling events?
- Are any site conditions or insights from prior sampling activities noted on field sampling sheets, such as "high turbidity" or "well pumped dry," that might point to localized well influences?
- Do the passive sampling results lead to the same site decisions as the historical data?
- If multiple passive samplers were used to profile a well, are the results from the samplers similar to each other? If not, do the active sampling results fall somewhere between the points? For long-screen wells or transmissive zones/lithological changes surrounding the saturated well screen, additional considerations or analysis may be needed.
- Were equivalent QA/QC methods employed for all methods being compared?
- If comparison of results is favorable, what other practical considerations for the different methods might be relevant to evaluate for the site (for example safety, cost/efficiency, equipment and staffing needs, sustainability, IDW management)?