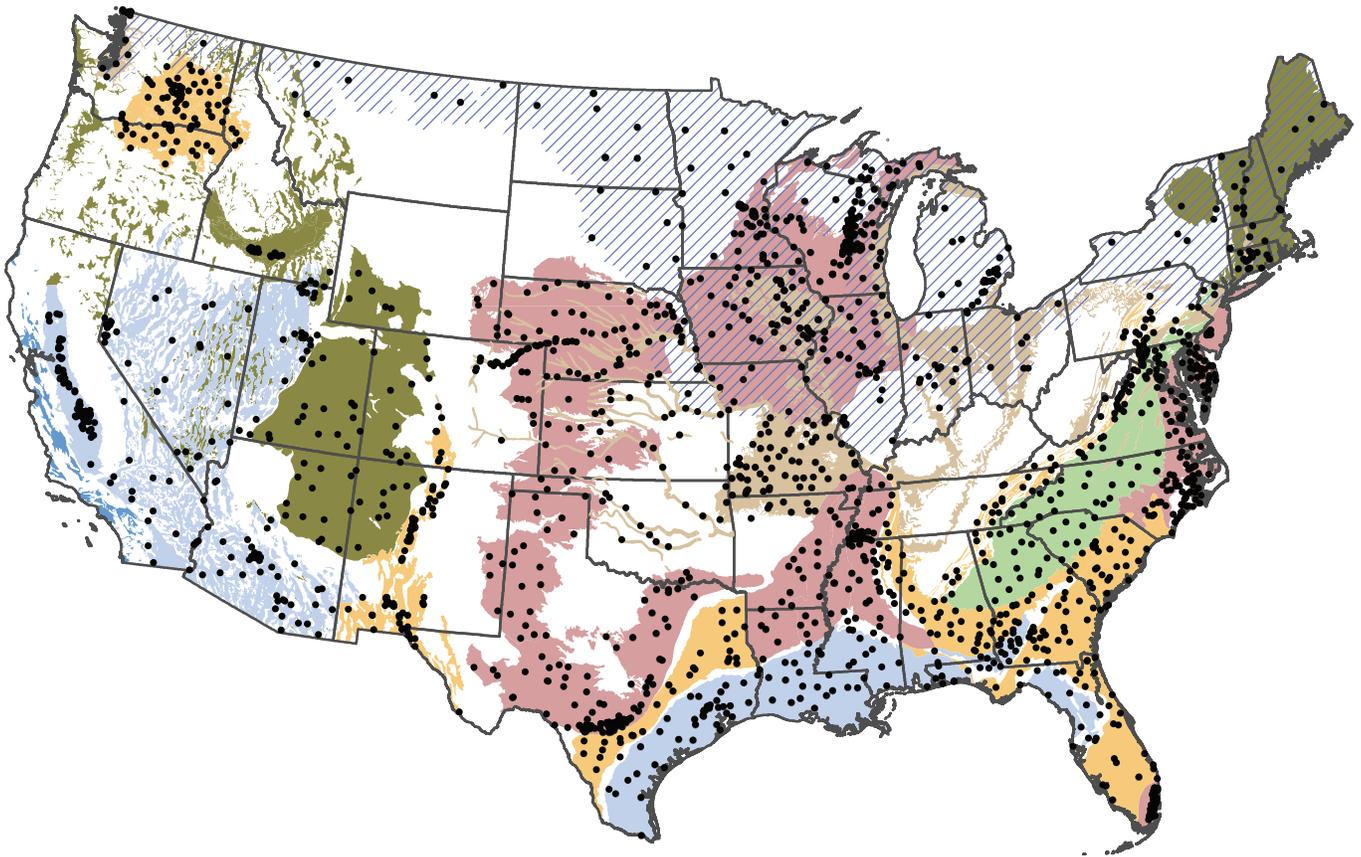


National Water-Quality Assessment Project

**Quality of Pesticide Data for Groundwater Analyzed for the
National Water-Quality Assessment Project, 2013–18**



Scientific Investigations Report 2020–5072

Quality of Pesticide Data for Groundwater Analyzed for the National Water-Quality Assessment Project, 2013–18

By Laura M. Bexfield, Kenneth Belitz, Mark W. Sandstrom, Delicia Beaty,
Laura Medalie, Bruce D. Lindsey, and Lisa H. Nowell

National Water-Quality Assessment Project

Scientific Investigations Report 2020–5072

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2020

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <https://store.usgs.gov>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Bexfield, L.M., Belitz, K., Sandstrom, M.W., Beaty, D., Medalie, L., Lindsey, B.D., and Nowell, L.H., 2020, Quality of pesticide data for groundwater analyzed for the National Water-Quality Assessment Project, 2013–18: U.S. Geological Survey Scientific Investigations Report 2020–5072, 35 p., <https://doi.org/10.3133/sir20205072>.

Associated data for this publication:

Bexfield, L.M., Sandstrom, M.W., and Beaty, D., 2020, Field, laboratory, and third-party data for assessment of the quality of pesticide results reported by the National Water Quality Laboratory for groundwater samples collected by the National Water-Quality Assessment Project, 2013–18: U.S. Geological Survey data release, <https://doi.org/10.5066/P90BFKA4>.

Contents

Abstract.....	1
Introduction.....	2
Purpose and Scope	3
Previous Data-Quality Assessments for Organic Compounds.....	3
Methods.....	4
Sample Collection and Preparation.....	4
Collection of Field Samples.....	4
Preparation of Laboratory Quality-Control Samples.....	5
Preparation of Third-Party Quality-Control Samples.....	5
Laboratory Analysis by Schedule 2437	7
Data Compilation.....	9
Data Preparation and Evaluation	9
Data for Field and Laboratory Blanks.....	10
Data for Third-Party Blind Blank Samples.....	13
Data for Laboratory Reagent Spikes, Field Matrix Spikes, and Third-Party Blind Spikes	14
Data for Field Replicates	19
Results of Data-Quality Assessment	19
Outcome of Laboratory Verification.....	19
Contamination Bias.....	19
Field Blanks.....	20
Laboratory Blanks.....	20
Third-Party Blind Blank Samples and Blind Spike Samples.....	22
Bias and Variability Indicated by Results for Spike Samples.....	23
Laboratory Reagent Spikes.....	23
Third-Party Blind Spike Samples	24
Field Matrix Spikes	25
Variability Indicated by Results for Field Replicates.....	26
Implications for Interpretation of Schedule 2437 Pesticide Results for Groundwater	29
Considerations Regarding Contamination Bias	29
Considerations Regarding Bias in Recovery.....	31
Considerations Regarding Variability.....	31
Summary.....	31
Acknowledgments.....	33
References Cited.....	33

Figures

1. Diagram showing principal aquifers and groundwater sites sampled across the United States in 2013-18 and contributing data to this analysis
2. Graphs showing reported detections and concentrations for groundwater samples, field blanks, and laboratory blanks plotted by analysis date for atrazine; metolachlor; and diketonitrile isoxaflutole

3. Graphs showing reported concentrations for groundwater samples, field blanks, and laboratory blanks plotted in the form of empirical distribution functions for atrazine in water year 2014 and fipronil sulfone in water year 2015.....	12
4. Graphs showing recovery values in laboratory reagent spikes, field matrix spikes, and third-party blind spikes plotted by analysis date for imazamox; aldicarb sulfoxide; and acetochlor sulfonic acid.....	15
5. Box plots showing comparison of recovery values in laboratory reagent spikes and field matrix spikes for naled and orthosulfamuron.....	18

Tables

1. Information about pesticide compounds included on the National Water Quality Laboratory schedule 2437, with human-health benchmarks and ranges of detection and laboratory reporting limits.....	https://doi.org/10.3133/sir20205072
2. Description of quality-control samples included in this evaluation and information they provide.....	6
3. Summary of detections of schedule 2437 pesticide compounds in groundwater, field blanks, and laboratory blanks, May 2013 through September 2018.....	https://doi.org/10.3133/sir20205072
4. Comparison of the 90-percent upper confidence limit for the 90th, 95th, and 99th percentiles of schedule 2437 pesticide compounds in laboratory blanks by water year, with minimum groundwater concentrations, maximum laboratory limits, and human-health benchmarks.....	https://doi.org/10.3133/sir20205072
5. Summary of false-positive and false-negative results for schedule 2437 pesticide compounds based on third-party blind spike samples, May 2013 through September 2018.....	https://doi.org/10.3133/sir20205072
6. Summary statistics for the recovery of schedule 2437 pesticide compounds in laboratory reagent spikes, May 2013 through September 2018.....	https://doi.org/10.3133/sir20205072
7. Summary statistics for the recovery of schedule 2437 pesticide compounds in third-party blind spike samples, May 2013 through September 2018.....	https://doi.org/10.3133/sir20205072
8. Summary statistics for the recovery of schedule 2437 pesticide compounds in field matrix spikes, May 2013 through September 2018.....	https://doi.org/10.3133/sir20205072
9. Estimated variability in detection of schedule 2437 pesticide compounds based on field replicates, May 2013 through September 2018.....	27
10. Estimated variability in concentration of schedule 2437 pesticide compounds based on field replicate samples, May 2013 through September 2018.....	28
11. Summary of results of data-quality assessment for schedule 2437 pesticide compounds based on all quality-control sample types, May 2013 through September 2018.....	https://doi.org/10.3133/sir20205072

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
Volume		
liter (L)	33.81402	ounce, fluid (fl. oz)
liter (L)	2.113	pint (pt)
liter (L)	1.057	quart (qt)
liter (L)	0.2642	gallon (gal)
liter (L)	61.02	cubic inch (in ³)
Mass		
gram (g)	0.03527	ounce, avoirdupois (oz)
kilogram (kg)	2.205	pound avoirdupois (lb)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$^{\circ}\text{F} = (1.8 \times ^{\circ}\text{C}) + 32.$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$^{\circ}\text{C} = (^{\circ}\text{F} - 32) / 1.8.$$

Datum

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Supplemental Information

Specific conductance is given in microsiemens per centimeter at 25 degrees Celsius ($\mu\text{S}/\text{cm}$ at 25 °C).

Concentrations of chemical constituents in water are given in nanograms per liter (ng/L).

Abbreviations

BBS	blind blank sample
BSS	blind spike sample
DL	detection limit
EDF	empirical distribution function
EPA	U.S. Environmental Protection Agency
FB	field blank
FMS	field matrix spike
FR	field replicate
GAMA	Groundwater Ambient Monitoring and Assessment
HBSL	Health-Based Screening Level
HHB	human-health benchmark
HHBP	Human-Health Benchmark for Pesticides
LRS	laboratory reagent spike
LB	laboratory blank
MCL	Maximum Contaminant Level
NAWQA	National Water-Quality Assessment
NWIS	National Water Information System
NWQL	National Water Quality Laboratory
QC	quality control
QSB	Quality Systems Branch
RL	reporting limit
RRL	raised reporting level
RSD	relative standard deviation
S2437	laboratory schedule 2437
SD	standard deviation
UCL	upper confidence limit
USGS	U.S. Geological Survey
VOC	volatile organic compound
VQC	value qualifier code

Quality of Pesticide Data for Groundwater Analyzed for the National Water-Quality Assessment Project, 2013–18

By Laura M. Bexfield, Kenneth Belitz, Mark W. Sandstrom, Delicia Beaty, Laura Medalie, Bruce D. Lindsey, and Lisa H. Nowell

Abstract

The National Water-Quality Assessment (NAWQA) Project of the U.S. Geological Survey (USGS) submitted nearly 1,900 samples collected from groundwater sites across the United States in 2013–18 for analysis of 225 pesticide compounds (pesticides and pesticide degradates, hereafter referred to as “pesticides”) by USGS National Water Quality Laboratory schedule 2437 (S2437). For the associated NAWQA study of pesticide occurrence and concentration in groundwater, and for other studies using pesticide results determined by S2437, it is necessary to assess the ability of reported results to meet data-quality requirements that will allow study objectives to be achieved. This assessment of the quality of S2437 results reported in 2013–18 examined data from field and laboratory quality-control samples, along with third-party performance assessment samples, to estimate bias and variability and to identify their potential sources, with an emphasis on implications for the interpretation of pesticide data for groundwater. Results indicate that measurements produced by the S2437 method for most pesticides have bias and variability that would be considered acceptable for many interpretative studies, which could therefore use the results without qualification or censoring. However, the reported data for a subset of pesticides have the potential for unacceptable contamination bias, high or low recovery bias, or high variability as a consequence of method performance and (or) nonlaboratory factors that could preclude their use for certain common objectives or could necessitate adjustment or qualification to meet those objectives.

Based on data for laboratory blanks, censoring of some detections for a subset of pesticides reported by the laboratory in environmental samples might be necessary or desirable to avoid an unacceptably high likelihood of a false-positive result caused by laboratory contamination. The 90-percent upper confidence limit for the 95th percentile of laboratory blank concentration equals or exceeds the minimum reported groundwater concentration in at least 1 water year for 28 pesticides. During at least 1 water year, this upper confidence limit exceeds the maximum laboratory detection limit for 17 pesticides and exceeds the maximum laboratory reporting limit for 3 pesticides (ametryn, atrazine, and

diazinon). The level of contamination indicated by this upper confidence limit should not substantially affect the suitability of reported environmental concentrations for any compound for comparison with corresponding human-health benchmarks.

Despite being subjected to the same laboratory processes as laboratory blanks, field blanks indicated little evidence of contamination bias. This observation could largely be the consequence of data-reporting practices, which utilize detections in laboratory blanks to censor results in associated field samples (including blanks and environmental samples) when relative concentrations indicate that a result could have a substantial contribution from laboratory contamination. Laboratory censoring appears likely to reduce the risk of false-positive results in environmental samples below the level that laboratory blank results alone would imply.

Whereas data available for third-party blind blank samples analyzed in 2018 indicate that only propoxur had any false-positive results, data for pesticides that were not spiked into blind spike samples analyzed in 2013–18 indicate that the false-positive rates for 31 pesticides exceeded 1 percent when considering only detections reported at concentrations greater than the maximum detection limit. Although about half of these pesticides lack substantial supporting evidence of contamination bias based on laboratory blank or field blank detections, indicating that spiking issues or degradation of parent compounds within the spiked samples might be a contributing factor to some false-positive results, these results indicate the need to closely examine detections reported for some pesticides in environmental samples analyzed during a similar period for possible contributions from contamination bias. Data for blind spike samples that were spiked at concentrations above the maximum reporting limit indicate that false-negative rates for eight pesticides exceed 10 percent; substantial low bias could affect results reported for these pesticides in environmental samples analyzed during a similar period.

Data for laboratory reagent spikes, which measure recovery of pesticides in blank water, show little evidence for unacceptable recovery bias for S2437 pesticides. However, field matrix spikes, which measure recovery of pesticides in environmental matrices, indicate that degradation and (or) matrix effects could result in moderate to substantial low bias for groundwater results for several pesticides. Low bias

could cause some reported concentrations to be categorized as being below a benchmark when the actual concentration in groundwater is greater than the benchmark. Occurrence and concentrations in groundwater could be substantially underrepresented for six pesticides with benchmarks (1H-1,2,4-triazole, asulam, bifenthrin, cis-permethrin, fenbutatin oxide, and naled) that have median recoveries between zero and 50 percent in field matrix spikes. Two compounds (didealkylatrazine and 2-hydroxy-6-ethylamino-4-amino-s-triazine) have median recoveries near or greater than 150 percent in field matrix spikes, indicating a substantial high bias. Plots of data for all spike types show clear changes in the typical recovery with time for some pesticides, which would require further examination for evaluation of temporal trends in environmental concentrations.

Data for laboratory reagent spikes indicate that nearly all S2437 pesticides have acceptable variability resulting from random measurement error. Only two compounds (fenbutatin oxide and naled) have F-pseudosigma values greater than 30 percent for recovery, which implies the potential for relatively high variability in reported concentrations and could affect comparison of concentrations to benchmarks and determination of whether concentrations for samples collected at separate locations or times are truly different with

a specified level of confidence. Data for third-party blind spike samples show relatively high variability for a greater number of pesticides, although these results likely reflect the influence of degradation and (or) differences in the magnitude and variability of concentrations used for blind spikes relative to laboratory reagent spikes. Detailed analysis of variability using field replicate data is possible for only 12 pesticides on S2437; low variability in analyte detection and concentration is indicated for most of these pesticides in groundwater.

Introduction

As part of its mission to characterize the quality of the Nation's water resources, the National Water-Quality Assessment (NAWQA) Project of the U.S. Geological Survey (USGS) submitted about 1,900 samples collected from groundwater sites across the United States in 2013–18 for analysis of pesticide compounds by USGS National Water Quality Laboratory (NWQL) method O-2437-15, under laboratory schedule 2437 (S2437) (Sandstrom and others, 2015). Groundwater samples were collected from wells and springs in 47 States and 27 principal aquifers (fig. 1),

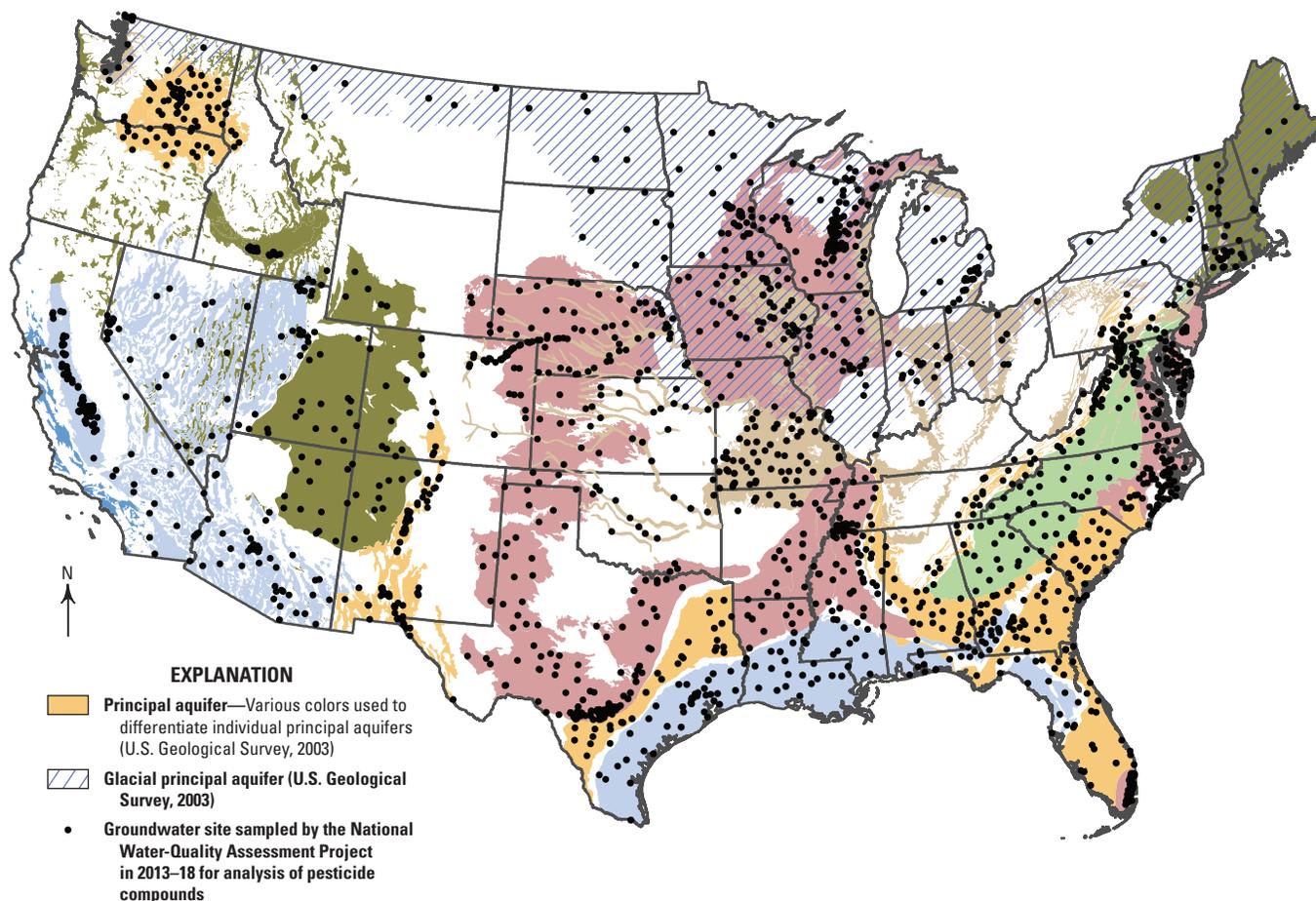


Figure 1. Principal aquifers and groundwater sites sampled across the United States in 2013–18 and contributing data to this analysis.

which are regionally extensive aquifers or aquifer systems that have the potential to be used for drinking water (USGS, 2003). A variety of well types was sampled, including public-supply, domestic-supply, and observation wells. NAWQA objectives for the collection of pesticide data for groundwater include characterization of compound occurrence and factors affecting their geographic distribution, along with evaluation of the potential relevance of detections of pesticide compounds to human health.

S2437 measures the occurrence and concentration of 225 compounds representing several pesticide types (herbicides, fungicides, and insecticides) and chemical classes (such as acetanilides, triazines, and acids), including many compounds that were not previously analyzed by the NWQL (table 1, available for download at <https://doi.org/10.3133/sir20205072>) (Sandstrom and others, 2015). The compounds include 109 pesticides and 116 pesticide degradates, many of which have human-health benchmarks (HHBs) for drinking water (table 1). (Hereafter, the term “pesticides” refers to pesticides and pesticide degradates unless otherwise indicated.) S2437 was made available for routine sample analysis in 2013 using a new liquid chromatography-tandem mass spectrometry method that had recently been developed by the NWQL to allow measurement of multiple pesticides at low detection limits (DLs), on the order of nanograms per liter, which supports thorough characterization of likely pesticide occurrence in the environment.

For studies conducted by NAWQA and others using S2437 pesticide data, it is necessary to assess the ability of data produced by sampling and analytical methods to meet data-quality requirements that will allow any individual study to achieve its objectives. Estimates of bias and variability are needed for a thorough quality assessment that characterizes measurement error and identifies any resulting limitations on the interpretation of the environmental data (Mueller and others, 2015). Bias is systematic error that is inherent in the measurement process and can be positive (causing reported sample concentrations to be consistently higher than concentrations actually present in the environment) or negative (causing reported sample concentrations to be consistently lower than concentrations actually present in the environment). Contamination bias is a common type of positive bias in water-quality samples that can be introduced while collecting, processing, transporting, and (or) analyzing a sample. Variability is random measurement error that is evidenced by disagreement between independent measurements.

Quality-control (QC) samples prepared in the field, in the laboratory, and by third parties provide data that can be used to assess bias and variability in the measurement process. In particular, blank samples prepared in all three settings can be used to evaluate contamination bias, spike samples prepared in all three settings can be used to evaluate overall bias, spike samples prepared in the laboratory and by third parties can be used to estimate variability, and replicate samples prepared in the field can contribute additional information regarding variability.

Purpose and Scope

The purpose of this report is to describe the quality of pesticide data for groundwater samples analyzed by S2437 for the National Water-Quality Assessment Project during 2013-18. The analysis objectives are to assess and document the quality of data by estimating (1) bias and (2) variability using data from field and laboratory QC samples, along with third-party performance assessment samples. When these samples indicate the potential for unacceptable bias or variability, likely contributing sources—laboratory method performance and (or) one or more nonlaboratory factors—are identified. Given results of the data-quality assessment, implications for interpretation of S2437 results for groundwater samples also are described.

Previous Data-Quality Assessments for Organic Compounds

Mueller and others (2015) provided guidance on the analysis and interpretation of field QC data (including for blank, spike, and replicate samples, which are described in detail in the “Sample Collection and Preparation” section of this report) and discussed studies of previous data-quality assessments of organic compounds by NAWQA. One such assessment by Martin (2002) examined the variability of pesticide detections and concentrations in field replicates (FRs) collected for surface water and groundwater in 1992–97 for analysis by laboratory schedules used by NAWQA at that time. Recoveries of laboratory reagent spikes (LRSs) and (or) field matrix spikes (FMSs) for pesticides were examined by Bexfield (2008), Martin and others (2009), and Martin and Eberle (2011) with the primary objective of determining implications for the analysis of temporal trends in NAWQA data for groundwater and (or) surface water. Bender and others (2011) analyzed laboratory blanks (LBs) and various types of blank samples collected in the field for volatile organic compounds (VOCs) to assess contamination bias and the potential need for censoring of surface-water and groundwater data collected by NAWQA, primarily because of detections in field blanks (FBs). Fram and others (2012) conducted a similar analysis of blank data for VOC samples collected for groundwater by another large USGS project, the California Groundwater Ambient Monitoring and Assessment Program Priority Basin Project (GAMA).

Additional studies have examined field, laboratory, and (or) third-party QC data to evaluate data quality for pesticides determined by S2437 and (or) older NWQL schedules used by NAWQA and (or) GAMA. Based on LBs, Medalie and others (2019) characterized positive bias from laboratory contamination for selected pesticides on laboratory schedules used by NAWQA in 2001–15, not including S2437; bias as indicated by false-positive and false-negative occurrence (defined in the “Sample Collection and Preparation” section of this report) determined from blind blank samples

(BSSs) and blind spike samples (BSSs) also was assessed. Fram and Stork (2019) examined FBs and LBs to quantify contamination bias for laboratory schedules used by GAMA in 2004–18, not including S2437. Martin and others (2017) analyzed results of a field study designed to evaluate bias and variability in pesticide results determined for surface water by S2437 in comparison with results determined by previous laboratory schedules used by NAWQA; QC samples evaluated for the Martin and others (2017) study included FMSs, LRSs, FBs, and FRs. Shoda and others (2018) described supplemental data-quality information provided by the laboratory when reporting pesticide results determined by S2437 and examined FMSs and LRSs analyzed in 2013–15 to characterize bias and variability.

Methods

Data compiled and analyzed for this quality assessment of pesticide results reported under S2437 included data for QC samples collected by NAWQA sampling crews in the field and for QC samples prepared by the NWQL and by a third party, the USGS Quality Systems Branch (QSB). Environmental data collected by NAWQA field crews from groundwater sites were compared with QC results to help determine and illustrate potential implications of the data-quality assessment for interpretation of groundwater results. Environmental and QC data were retrieved from appropriate databases for the period of May 2013, when NAWQA collection of groundwater samples for analysis by S2437 began, through the end of water year 2018 (hereafter, unless otherwise stated, a specified year will refer to water year, which is defined as the 12-month period from October 1 of any given year through September 30 of the following year and is designated by the calendar year in which it ends). Statistical and graphical summaries of QC and environmental sample results were used to identify potential data quality issues for further analysis. Because data extended over several years, changes in bias and variability through time could be graphically illustrated.

Sample Collection and Preparation

This section of the report describes the collection of field samples and the preparation of laboratory and third-party QC samples.

Collection of Field Samples

Field protocols used for equipment cleaning, sample collection, and sample processing are described in detail in chapters A3, A4, and A5 of the USGS National Field Manual (USGS, variously dated); only brief descriptions of sample collection and processing are provided here. Environmental samples are collected as raw groundwater samples at or near the wellhead or, for springs, as close to the groundwater

source as possible. Field properties consisting of water temperature, specific conductance, pH, dissolved oxygen, and alkalinity are measured on location. Samples are collected after three casing volumes of groundwater have been purged from the well and field properties have stabilized or, for springs, after determining that field properties are stable. Samples for analysis of pesticides by S2437 are collected using a Teflon sampling line and filtered through a 0.7-micron baked glass-fiber filter into 20-milliliter (mL) amber glass bottles; samples are shipped on ice overnight to the NWQL in Denver, Colorado, for analysis. Dedicated pumps are used for sample collection when present, such as for public-supply and domestic wells. Portable test pumps are used for sample collection when a dedicated pump is not present, which is typical of most observation wells sampled by NAWQA.

Blank samples collected by NAWQA field crews for pesticides consist of FBs, which are collected at a sampling site in a manner comparable to collection of an environmental sample. They are intended to evaluate the potential for the various aspects of sample collection, field processing, preservation, transportation, and laboratory handling to be sources of contamination (that is, positive bias) (table 2) (Mueller and others, 2015). FBs for pesticides are assigned at the quantity of 1 FB for every 15 wells sampled or 1 FB for each team sampling a given network of wells, whichever results in a greater number (Arnold and others, 2018). Nitrogen-purged organic blank water prepared to be free of measurable concentrations of pesticides is obtained from the NWQL for use in collection of FBs. Certificates of analysis are available from the NWQL for all blank-water lots, and examination of these certificates did not indicate contamination of any lots. When an FB is collected for a well without a dedicated pump, the same portable test pump that is used to collect the environmental sample from the well is also used to pump blank water through any other appropriate sampling equipment and the filter and into the sample bottle. When an FB is collected for a well with a dedicated pump, a valveless piston metering pump or peristaltic pump with wetted parts constructed of materials appropriate for collection of samples for organic compounds is used to pump blank water through any appropriate sampling equipment and the filter and into the sample bottle.

Other QC samples collected by NAWQA field crews for pesticides consist of FMSs, collected at the quantity of 1 FMS for every 30 wells sampled, and FRs, collected at the quantity of 1 FR for every 15 wells sampled (Arnold and others, 2018). An FMS for groundwater is prepared in the field by fortifying a sample of environmental water, collected sequentially to the main environmental sample, with 100 microliters (μL) of spike solution obtained from the NWQL (described in the “Preparation of Laboratory Quality-Control Samples” section); certificates of analysis showing the solution concentrations of all compounds are available from the NWQL for all spike lots. Each spiked sample is accompanied by an unspiked environmental sample to correct the spike data for background concentrations in determination of recovery, defined as the concentration measured in the spike expressed as a percentage

of the known concentration that was added to the sample (Mueller and others, 2015). FMSs are used to estimate any positive or negative bias that might result from a combination of method performance (laboratory environment and analysis), effects of the sample matrix, and (or) degradation of a compound between sample collection and analysis (table 2). An FR is a sample of environmental water that, in this case, is collected sequentially to the main environmental sample with which it is paired. An FR is collected, processed, and analyzed in a manner allowing it to be considered essentially identical to the main sample. Comparison of results for the main sample and its paired FR allows estimation of the variability (random measurement error) of analytical results; a sequential replicate also can be affected by any temporal change in the sampled medium between collection of the paired samples (table 2) (Mueller and others, 2015), but such change typically is expected to be negligible for groundwater.

Preparation of Laboratory Quality-Control Samples

Preparation of LBs and LRSs is described in detail in Sandstrom and others (2015); only brief descriptions of sample preparation are provided here. LBs are prepared using reagent water and are included among routine batch and instrument QC samples. At least 1 batch blank (a 10-mL sample in a 20-mL analytical vial, typically analyzed after the first set of 15 environmental samples) and several instrument blanks (a 1-mL sample in a 2-mL analytical vial, typically analyzed before and after each set of 15 environmental samples) are included in each analytical batch of 75 environmental samples and associated QC samples. Together, these two types of LBs provide information about positive bias resulting from contamination during sample preparation and analysis in the laboratory (table 2), such as from use of insufficiently cleaned glassware, and (or) from carryover of analytes among samples resulting from insufficient rinsing of instrument equipment between samples. LRSs are prepared by adding 100 μ L of spike solution to about 10 mL of reagent water. Spike solutions are prepared by the NWQL by using analytical reference standards, typically to result in nominal concentrations of 250 nanograms per liter (ng/L) for each analyte in a sample (the same spike solutions are used for FMSs). One or two LRSs typically are included in an analytical batch. Recovery values for LRSs provide information about bias (positive or negative) and variability resulting from method performance (laboratory environment and analysis) (table 2).

Preparation of Third-Party Quality-Control Samples

The QSB is independent from the NWQL and submits third-party BBSs and BSSs to the NWQL for analysis in order to assess laboratory performance. These samples use reagent

water but otherwise are designed to be indistinguishable from environmental samples submitted by NWQL customers. Results for QSB samples may be indicative of laboratory method performance for environmental samples submitted during the same time period because they undergo the same laboratory processes. Although QSB samples do not mimic an environmental matrix or handling of field samples, which can also affect results reported by the laboratory, results for BSSs do have the potential to reflect the effects of degradation of spiked compounds in the few to several days between the time of sample preparation and laboratory analysis (degradation of compounds within spike solutions prior to spiking is expected to be minimal because of QSB storage protocols for these solutions). In contrast, degradation is believed to have minimal effect on results for LRSs because LRSs typically are analyzed within a day or so of preparation.

BBSs consist of reagent water and are used to estimate the false-positive occurrence for individual compounds (that is, the occurrence of a reported detection when the compound is not actually present in the sample). False-positive occurrence generally is considered a measure of positive bias resulting from contamination introduced during laboratory processes (table 2), although false-positive results can occur because of misidentification resulting from incorrect application of qualitative identification criteria (described in the “Laboratory Analysis by Schedule 2437” section of this report).

BSSs consist of reagent water spiked with a mix of a subset of compounds on the laboratory schedule. Because no laboratory other than the NWQL is known to analyze for all of the pesticides on S2437, or to analyze for overlapping compounds at similarly low concentrations, meaningful independent verification of the spike solutions used by the QSB is not possible. Therefore, new spike solutions are used to prepare high-concentration BSSs for analysis by the NWQL to provide semiquantitative verification of compounds and concentrations in the mixes prior to use. Throughout the period of use of a mix of spike solutions, results reported by the NWQL for BSSs are reviewed for evidence of degradation or other issues that can result in individual solutions or entire mixes being replaced and associated data being rejected as unrepresentative of laboratory performance. Results typically are rejected only when essentially all samples using a certain solution and (or) mix show likely degradation or errors in the compounds and (or) concentrations included.

Similar to BBSs, BSSs can be used to estimate the false-positive occurrence for an individual compound when that compound was not in the mix spiked into the sample, although with somewhat lower reliability than can be achieved using BBSs because of the potential for the compound to be present in the sample as a result of a spiking error or, for degradate compounds, degradation of the parent compound. As the spiked concentration of a parent compound increases, the potential for degradation to cause a high enough concentration of the degradate to produce a false-positive result can likely increase. Also, misidentification of compounds or interferences

6 Quality of Pesticide Data for Groundwater Analyzed for the National Water-Quality Assessment Project, 2013–18

Table 2. Description of quality-control samples included in this evaluation and information they provide (modified from Mueller and others, 2015).

[T, targeted source; X, additional (nontargeted, but unavoidable) source; –, not a potential source]

Blank samples				
Potential source of contamination	Type of blank sample			
	Field	Laboratory batch blank	Laboratory instrument blank	Third-party blind blank
Field sources				
Field environment				
Air, rain, dust, fumes	T	–	–	–
Sample-collection personnel				
Dirty hands, personal care products	T	–	–	–
Sample collection				
Samplers, pumps, tubing	T	–	–	–
Sample processing				
Splitters, filters, chambers	T	–	–	–
Sample bottles or vials	T	–	–	–
Sample preservation	T	–	–	–
Equipment cleaning				
Soap, inadequate rinsing, carryover	T	–	–	–
Transport to and from the field site				
Field vehicles, coolers	T	–	–	–
Shipping to laboratory				
Coolers, commercial carriers	T	–	–	–
Laboratory sources				
Laboratory environment and analysis	T	T	T	T
Carryover among samples in a batch	T	–	T	T
Other sources				
Water used to make the blank ¹	X	X	X	X

Spike samples			
Potential source of bias	Field matrix spike	Laboratory reagent spike ²	Third-party blind spike ²
Field sources			
Field environment			
Water matrix interference	T	–	–
Shipping to laboratory			
Analyte degradation	T	–	X
Laboratory sources			
Laboratory environment and analysis	T	T	T

Table 2. Description of quality-control samples included in this evaluation and information they provide (modified from Mueller and others, 2015).—Continued

[T, targeted source; X, additional (nontargeted, but unavoidable) source; –, not a potential source]

Replicate samples	
Potential source of variability	Sequential replicate
Field sources	
Temporal change in sampled medium ³	X
Laboratory sources	
Laboratory environment and analysis	T

¹Although certified as appropriate for preparing blanks, there is a possibility that blank water can be contaminated during shipment or storage before use.²Laboratory reagent spikes and (in the absence of degradation) third-party blind spikes also can be used to assess variability (random measurement error).³Expected to be negligible for groundwater.

among compounds could contribute to some false-positive results, although interference would only be possible for compounds with similar retention times and characteristic ion pairs, which is not common among S2437 pesticides. BSSs that contain a certain compound can be used to estimate the false-negative occurrence for that compound (that is, the lack of a reported detection when the compound is actually present in the sample), which is an indicator of negative bias resulting from laboratory processes. However, degradation of some parent compounds between spiking and sample analysis could result in some fraction of the reported false-negative occurrence for those parent compounds being the result of factors other than laboratory performance.

Recovery values for compounds included in BSSs provide information about bias (positive or negative) and variability resulting from sample processing and analysis (table 2), although degradation of some parent compounds could result in lower recovery values for those parent compounds and higher recovery values for the degradate compounds than the laboratory would achieve if no degradation had occurred. Also, because compounds typically are spiked at varying concentrations near (a factor of 2–3 times) the laboratory reporting limit (RL) in effect at the time of spike preparation, which for most compounds is lower than the nominal concentration of 250 ng/L for LRSs, the false-negative occurrence based on these spiked samples indicates the frequency at which a compound might not be positively identified and reported in a sample if present at a concentration near the RL. Similarly, the resulting recovery values generally indicate bias and variability at relatively low concentrations, although samples are occasionally spiked at substantially higher concentrations.

Laboratory Analysis by Schedule 2437

Sandstrom and others (2015) detailed the analytical procedure used to determine pesticide compounds in filtered water samples by S2437, and Shoda and others (2018) described supplemental data-quality information provided by

the laboratory when reporting pesticide results determined by S2437; only brief descriptions of aspects of analytical and reporting procedures relevant to data-quality assessment are provided here. Through 2018, the S2437 analytical method for pesticides allowed for the determination of 225 compounds, including 109 pesticides and 116 pesticide degradates (table 1), that demonstrated acceptable laboratory performance to be retained on the schedule. This method employs direct injection of a 100- μ L aliquot of filtered sample onto a liquid chromatographic column for separation and subsequent determination by a tandem mass spectrometer operated in positive electrospray ionization mode for one group of pesticides and in negative electrospray ionization mode for a second group (Sandstrom and others, 2015). Analytes are identified on the basis of agreement between known standards with respect to chromatographic retention time and the ratios of characteristic quantification and qualifier ions, with consideration of the signal to noise ratio for the chromatographic peaks. A solution of 21 isotope-labeled (enriched) pesticides used as surrogates (table 1) is added to each sample and is used to quantify the pesticides determined in the method, compensating for matrix effects and other variables in the determined concentrations and thereby improving quantitation. Most pesticides included in the method generally can be determined at concentrations as low as 1–250 ng/L. The typical sample sequence for environmental and quality-control samples in an analytical batch is discussed further in Sandstrom and others (2015).

DLs for S2437 pesticides were initially calculated by using the U.S. Environmental Protection Agency (EPA) test procedure guidelines for analysis of pollutants (40 CFR, part 136) (Sandstrom and others, 2015). The DL is defined as the minimum concentration of a substance that can be measured and reported with 99-percent confidence that the analyte concentration is distinguishable from method blank results (resulting in a false-positive risk of no more than 1 percent). To reduce false-negative risk to no more than 1 percent, initial RLs (used in reporting nondetections) were set at two times the DL and were subsequently adjusted as

needed on the basis of performance considerations. Beginning in 2016, DLs and RLs for most pesticides were determined by using selected components of ASTM International's Standard Practice D6091–07 (ASTM International, 2007) and supporting DQCALC software (Standard Practice D7510-10; ASTM International, 2010), as described by USGS (2015). Table 1 lists the DLs and RLs in effect before and after January 1, 2016, in addition to the maximum DL and RL for the entire 2013–18 time period.

The maximum DL and RL are lower than the HHB for nearly all of the 119 pesticides on S2437 that have an HHB, providing confidence that the pesticides typically will be identified and quantified if they are present in the environment at concentrations that might be of concern for human health. The HHBs (table 1) are a set of health-based comparison thresholds that include EPA Maximum Contaminant Levels (MCLs) (EPA, 2018), Cancer and Noncancer Health-Based Screening Levels (HBSLs) (Norman and others, 2018), and Carcinogenic and Chronic Noncancer Human-Health Benchmarks for Pesticides (HHBPs) (EPA, 2017); the lowest available HHB for each pesticide is included in table 1. No pesticides have a maximum DL greater than the HHB, and only three pesticides (cyanazine, fentin, and oxyfluorfen) have a maximum RL greater than the HHB. In addition, the maximum DL and RL are lower than 10 percent of the HHB for 114 of 119 pesticides. The maximum DL is greater than 10 percent of the HHB for bromoxynil, cyanazine, fentin, and oxyfluorfen; the maximum RL is greater than 10 percent of the HHB for these pesticides plus terbufos. These few pesticides are less likely than others to be identified and quantified (that is, they are more likely to be susceptible to false-negative risk) when they are present in the environment at concentrations near the HHB.

For S2437 pesticides, all concentrations that meet identification criteria and relevant quality-control criteria are reported down to the lowest reportable value (table 1), which is typically 10 percent of the DL (or about 5 percent of the RL). For environmental samples (which for the purposes of NWQL data reporting include any sample other than a laboratory QC sample), when an analyte is not detected or does not meet qualitative criteria and is below the RL, it generally is reported as less than (<) the RL; for LBs, a nondetection generally is reported as zero. However, results for environmental samples and LBs are sometimes reported as less than a concentration higher than the RL, known as a raised reporting level (RRL). An RRL can be reported for a variety of reasons, typically documented in a laboratory comment and (or) a value qualifier code (VQC) accompanying the result, and can range up to orders of magnitude higher than the standard RL. As described by Shoda and others (2018), reasons for RRLs can be broadly classified into one category in which there is uncertainty in the identification of the pesticide and a second category in which there is uncertainty in the quantification of the pesticide.

One situation that can cause an environmental result to be reported with an RRL and (or) other data qualifiers is detection of contamination in one or more LBs (USGS, 2011).

Depending on the relative concentrations of the compound in the LB(s) and environmental sample(s), set-by-set censoring may be applied to the 15 environmental samples (sample set) analyzed directly before and after the blank (Shoda and others, 2018; Medalie and others, 2019). If the concentration measured in the environmental sample is more than 3 times but less than or equal to 10 times the concentration in the associated LBs, the environmental sample result is reported with a VQC of “v” (contamination) and a remark code of “E.” If the measured environmental concentration is less than three times the concentration in the associated LBs and is greater than the RL, the environmental sample result is reported with a VQC of “v” and a remark code of “<,” which effectively turns the environmental sample result into a nondetection reported at an RRL set equal to the measured environmental sample concentration. If the environmental concentration is more than 10 times the concentration in the associated LBs, no qualifiers are added to the reported sampled concentration. If the measured environmental concentration is less than 3 times the concentration in the associated LBs and is less than or equal to the RL, the environmental result is reported as less than the RL, with no additional qualifiers. This set-by-set censoring approach differs for S2437 relative to previous NWQL pesticide schedules in that detections in instrument blanks, rather than just in batch blanks, will initiate censoring, likely resulting in censoring being more common.

On May 21, 2015, the NWQL began consistent application of censoring rules for S2437 pesticide results for environmental samples related to interferences in associated LBs. Censoring was applied when interferences in the associated LBs were considered chronic (in more than 50 percent of LBs in the analytical batch) and prevented qualitative identification of the analyte in the environmental samples. Following the same general approach used in set-by-set censoring for LB contamination, as described in the preceding paragraph, censoring rules for interference are concentration dependent and can result in detections at low concentrations being reported as less than the RL (that is, as nondetections) with no additional qualifiers. Depending on the response of the analyte or the interference in the environmental sample relative to that determined for the LBs, the environmental sample result is reported with a VQC of “i” and a remark code of “E” to indicate the possibility of positive bias related to similar levels of interference. Because environmental results reported prior to May 21, 2015, were not subsequently reevaluated under these censoring rules for interference, it is possible that some groundwater detections reported during that time would have been reported as nondetections under the new rules. Determination of any potential effects on detection rates for individual pesticides in groundwater before versus after this date is not possible based on the information reported by the NWQL.

Results for 19 pesticides analyzed under S2437 have been routinely reported with qualifying codes since method implementation because of known issues with bias or variability by this method in 1 or more tested matrices

(Sandstrom and others, 2015; Shoda and others, 2018). These 19 pesticides, identified with a validation quality code of “E” in table 1, have all quantified results reported with an “E” remark code and a VQC code of “m,” meaning that results for the compound are highly variable. As a result of periodic review by the NWQL of method performance for all S2437 pesticides, an additional 15 pesticides began to be reported with these qualifying codes as of January 2016; these 15 pesticides are identified with a validation quality code of “E-2016” in table 1.

Data Compilation

Data for S2437 pesticides for NAWQA groundwater and field QC samples collected and analyzed between May 2013 and September 2018 were retrieved from the USGS National Water Information System (NWIS) database (USGS, 2019) in May 2019; therefore, data in the current report reflect results available in NWIS at that time for samples known to have been collected for the NAWQA project. Data for groundwater and field QC samples collected between May 2013 and December 2016 have been published previously in Arnold and others (2016, 2017, 2018, 2020); data for FMSs and FRs collected at groundwater sites between May 2013 and September 2015 for NAWQA also were published in Shoda and others (2017a, 2017b). However, in order to fully document data preparation and evaluation, data for all groundwater and field QC samples used for the current study, covering the period of May 2013 through September 2018, are provided in Bexfield and others (2020). For the current study, results for four pesticides—famoxadone (parameter code 67609), methyl parathion (parameter code 65089), phosmet (parameter code 65101), and pymetrozine (parameter code 68680)—that were eliminated from S2437 between 2013 and 2018 because of poor laboratory performance were excluded. Null values, typically resulting from an analytical issue, also were excluded. Results for groundwater and QC samples that arrived warm at the laboratory and (or) were extracted or analyzed past the holding time (14 days) were removed from the dataset. Only those QC samples with sufficient metadata to make them usable for their intended purpose (for example, type of blank for FBs and lot number of spike solution for FMSs) were retained. Samples with results providing clear evidence of some type of sampling or analytical error were excluded. Examples of such samples include a sample coded as an environmental sample with detections of most pesticides at concentrations indicating that the sample had almost certainly been spiked, a sample coded as an FMS that had few or no detections, and a sample with dilution codes for multiple pesticides that were confirmed to represent reporting errors.

Data for S2437 LBs and LRSs analyzed in 2013 through 2018 were retrieved in June 2019 and August 2019, respectively, from the NWQL Laboratory Information Management System; these data are not reported to NWIS. The retrieval of LBs included both batch blanks and instrument

blanks. Data for LRSs analyzed in water years 2013–15 were previously published in Shoda and others (2017b), and data for LRSs analyzed in water years 2016–17 were previously published in Wieben (2019). Data for LRSs analyzed in water year 2018 and for all LBs included in this study are provided in Bexfield and others (2020). LBs and LRSs associated with a type of environmental water sampler that was not used for groundwater sample collection were excluded. Null values, typically resulting from an analytical issue, and any duplicate values also were excluded. In addition, results for one LRS that were unusually high for most pesticides and were consistent with values having been mistakenly reported as concentrations rather than recoveries were removed from the dataset.

Data for third-party BBSs and BSSs analyzed by S2437 in 2013 through 2018 were retrieved in July 2019 from the QSB database. BBSs were submitted for analysis by S2437 starting only in October 2017. All BBS and BSS data included in the current study are provided in Bexfield and others (2020). Results reported as null values, typically resulting from an analytical issue, were excluded. Results for samples that arrived warm at the laboratory and (or) were extracted or analyzed past the holding time (14 days) were removed from the dataset. Results for BSSs were reviewed for evidence of a spiking error (that is, inclusion of a pesticide in a spike solution in which it was supposed to be excluded or exclusion of a pesticide from a spike solution in which it was supposed to be included) or degradation of a pesticide included in the spike solution that substantially reduced its concentration and (or) introduced one or more degradate compounds that should not have been present. Results believed to have been affected by these issues such that they were no longer representative of laboratory performance were excluded, although some remaining results might be affected by these issues to an undetermined degree, as discussed in the “Results of Data-Quality Assessment” section of this report.

Data Preparation and Evaluation

After environmental and QC data were compiled and incomplete or erroneous results were excluded as described in the “Data Compilation” section of this report, additional data preparation was required to allow evaluation of data quality. As detailed in this section, approaches to data preparation and evaluation differed by sample type and data objective. This section includes descriptions of plots that are provided in Bexfield and others (2020) as a means to visualize and further examine the data that are published in the tables of that data release and used in the current data-quality assessment; example plots are included in the figures of this report to demonstrate important concepts and support certain conclusions. Data quality objectives as specified here are intended as helpful criteria for evaluating the performance of sampling and analytical procedures and for deciding when certain results might need to be qualified or censored; they are not intended as strict criteria indicating the need for data rejection.

Data for Field and Laboratory Blanks

Data for FBs and LBs were evaluated for contamination bias for individual pesticides by using statistical summaries supplemented with graphical analysis. Specifically, basic statistical summaries were calculated by pesticide for FBs and LBs for the entire 2013–18 period to characterize the general frequency and magnitude of contamination in blank samples (table 3, available for download at <https://doi.org/10.3133/sir20205072>). These summaries include the total number of detections reported by the NWQL, the maximum reported concentration, the frequency of detection reported at any concentration, and the frequency of detection reported at a concentration at or above the maximum DL or RL listed in table 1. Maximum reported blank concentrations for each pesticide were also expressed as a fraction of the corresponding HHB to evaluate whether potential contamination bias is small enough that reported groundwater concentrations are suitable for use in determining exceedances of the HHB and (or) 10 percent of the HHB.

In addition, reported detections and concentrations in FBs, LBs, and groundwater samples were plotted together for individual pesticides by analysis date to compare the frequency, timing, and magnitude of detections among these sample types. Figure 2 shows results for atrazine, metolachlor, and diketonitrile isoxaflutole during the 2013–18 time period; graphs for other pesticides are available in Bexfield and others (2020, PlotGroup1). Detections and nondetections are represented using different symbols (fig. 2). Detections are plotted as solid circles at the reported concentration. Nondetections are plotted as open circles at the standard RL in effect at the time of analysis (identified on each graph) or, if applicable, at the RRL specified for an individual sample; this same plotting approach is used in figure 2 even for LBs, which typically have nondetections reported at a concentration of zero (values of zero are retained for use in other plots and for data analysis). Note that samples are not evenly distributed throughout each year because nearly all groundwater samples and FBs are collected in May through September, and more than half of LBs are analyzed in May through August in correspondence with increased environmental sample submissions. These temporal plots also were examined for detections reported for groundwater sites across multiple regions of the United States over short time periods of a few days or weeks (with or without detections in FBs or LBs during similar periods) for pesticides that otherwise were not commonly detected. This pattern of observing a large proportion of detections for an individual pesticide over a very short time period across multiple hydrogeologic and land-use settings was judged to be unlikely to occur under environmental conditions. Therefore, laboratory verification was requested for these detections; findings are discussed in the “Outcome of Laboratory Verification” section of this report.

In addition, empirical distribution functions (EDFs) (Helsel, 2005) illustrating the percentiles of observations for each sample type were created for selected pesticides

and water years to examine the distributions of FB and LB concentrations relative to groundwater concentrations for samples analyzed during that year. Figure 3 shows graphs comparing EDFs among these sample types for atrazine in water year 2014 and fipronil sulfone in 2015; graphs for other pesticides and water years are available in Bexfield and others (2020, PlotGroup2). The vertical axis shows percent, and the horizontal axis shows reported sample values (concentration). Sample values are plotted in sequence from low to high and are connected by lines. The lines begin at the percentile of the lowest detection among all samples (detections and nondetections) in the dataset; as percentiles increase, the lines reflect the plotting positions for both detections and any higher ranking nondetections (assignment of ranks is discussed in more detail later in this section). The plotting position for a given sample on the vertical axis gives an estimate of the percentage of results in the dataset having a value that is lower than the value for that sample. For the example in figure 3A, for any given percentage of sample results from 2014, the concentration in LBs is more than an order of magnitude lower than the concentration in groundwater samples.

The potential bias that could be introduced to groundwater samples as a result of contamination was estimated by constructing an upper confidence limit (UCL) for selected high percentiles of concentrations in blank samples. To estimate potential bias resulting from contamination at the laboratory, the 90-percent UCL was calculated for the 90th, 95th, and 99th percentiles of concentration for each pesticide that had been detected in LBs by using a binomial probability method based on one-sided, nonparametric confidence limits using the Jeffreys interval (Brown and others, 2001; Belitz and others, 2010; Fram and others, 2012). The calculations used in this approach, which ranks blank values from low to high and determines the rank of the blank corresponding to the specified UCL, are detailed in Fram and others (2012). The concentration of the blank at the rank corresponding to the specified UCL is the maximum contamination expected in the specified percentage of samples (Bender and others, 2011; Mueller and others, 2015). Therefore, the 90-percent UCL for the 95th-percentile concentration in LBs is the maximum concentration expected, with 90-percent confidence, in 95 percent of all samples; in other words, laboratory contamination is estimated, with at least 90-percent confidence, to exceed the 90-percent UCL for the 95th-percentile concentration of LBs in no more than 5 percent of all samples (Bender and others, 2011). Among others, Bender and others (2011), Fram and others (2012), and Bexfield and others (2019) have calculated UCLs for the 90th and (or) 95th percentiles of contamination in FBs and (or) LBs for application to interpretation of environmental data for organic compounds. Fram and others (2012) and Bexfield and others (2019) applied UCLs as censoring levels for some compounds, concluding that detections in environmental samples with concentrations lower than or equal to the UCL had an unacceptably high probability of resulting from contamination by the processes that affected the blanks. Fram

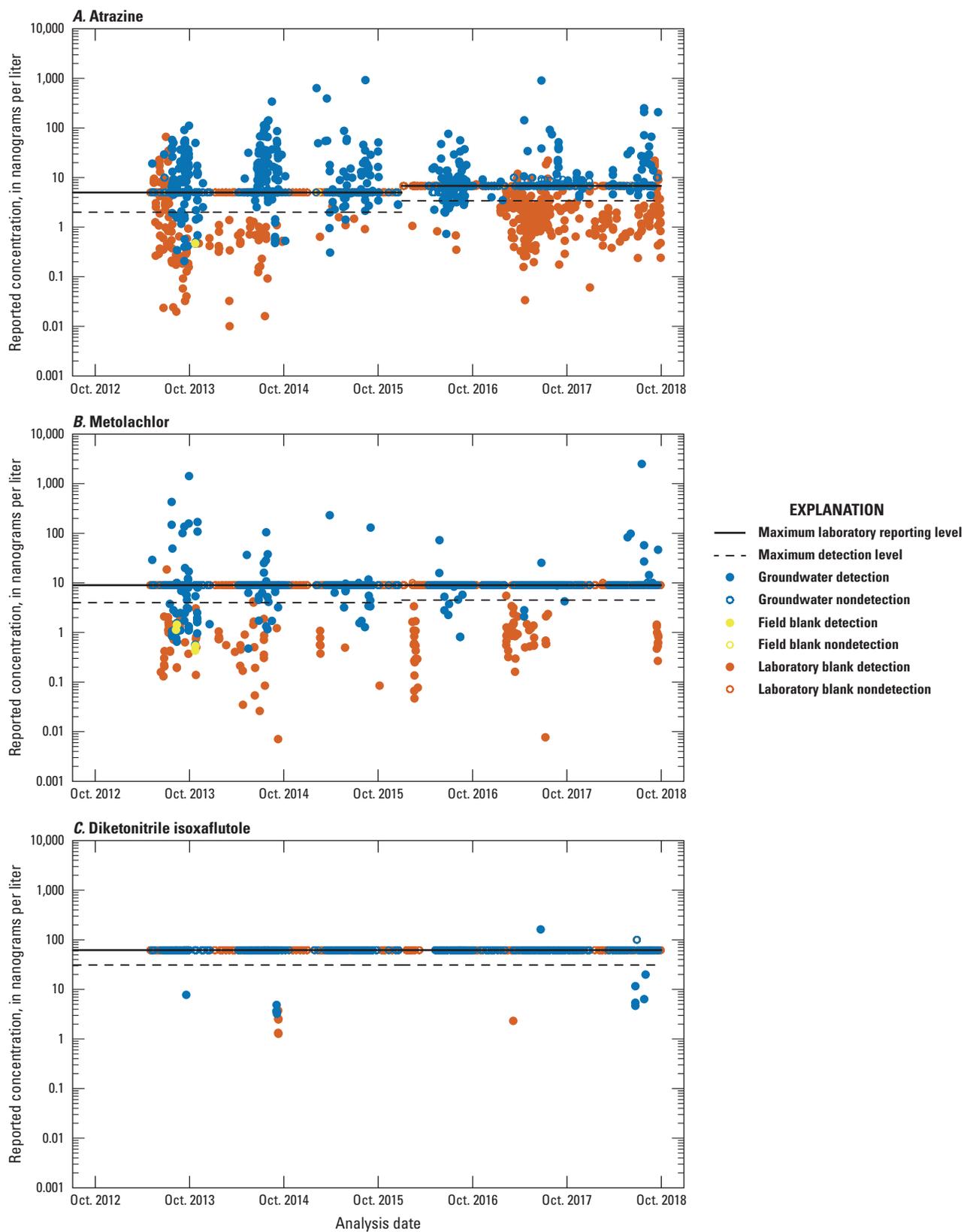


Figure 2. Reported detections and concentrations for groundwater samples, field blanks, and laboratory blanks plotted by analysis date for *A*, atrazine; *B*, metolachlor; and *C*, diketonitrile isoxaflutole.

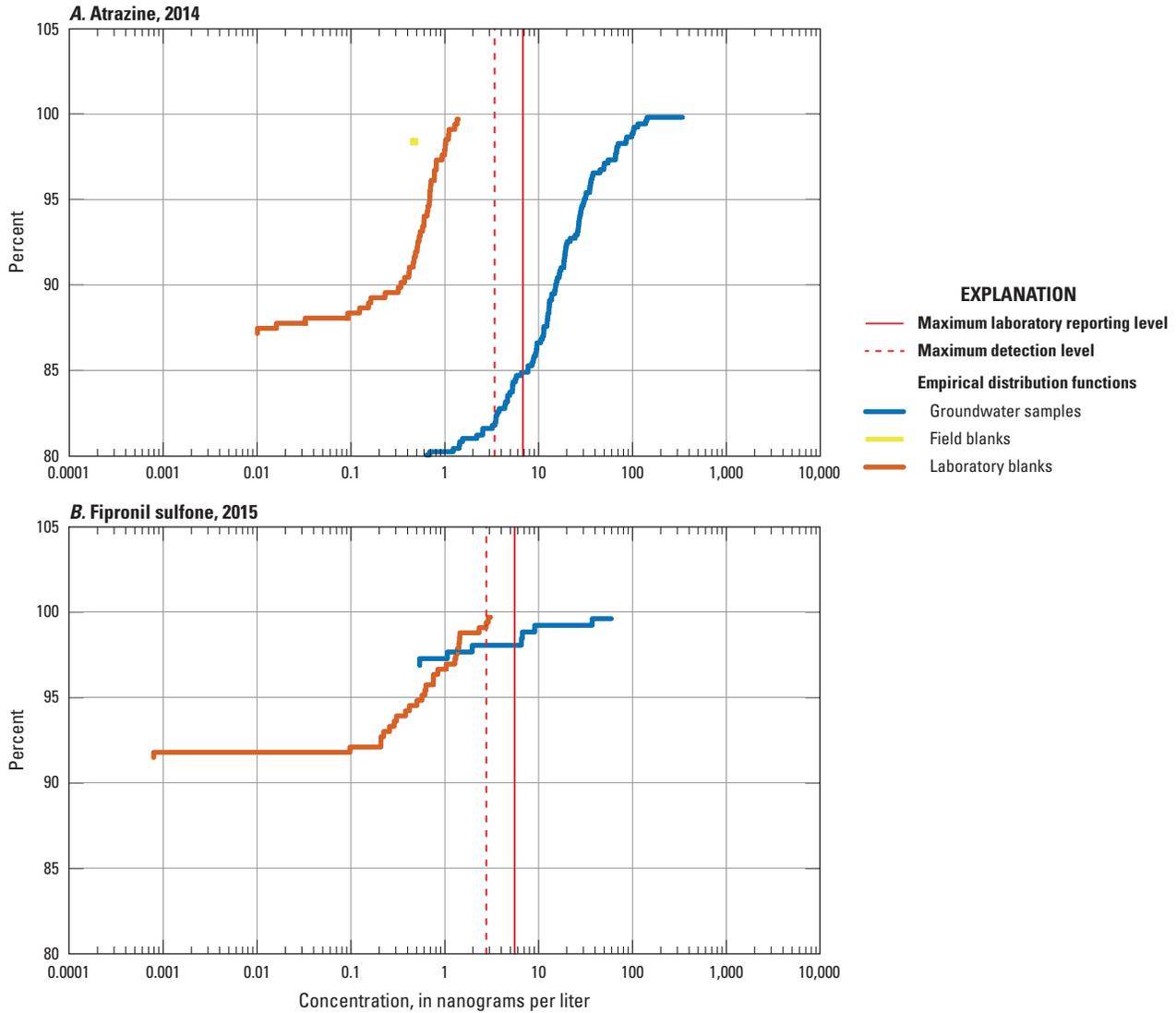


Figure 3. Reported concentrations for groundwater samples, field blanks, and laboratory blanks plotted in the form of empirical distribution functions for *A*, atrazine in water year 2014 and *B*, fipronil sulfone in water year 2015.

and Stork (2019) have specifically examined potential use of the 99th percentile of blank concentrations as a censoring level for individual pesticides. For the current study, UCLs were calculated for LBs by water year because initial graphical analysis indicated that water year (which typically reflects seasonal differences in intensity of sample collection) was a logical grouping to capture temporal variability in detection frequency and concentration for some pesticides in LBs, as discussed in the “Results of Data-Quality Assessment” section of this report. As also discussed in that section, UCLs were not calculated to estimate potential bias from field contamination because of the small number of detections reported for individual pesticides in FBs.

The percentiles used in EDFs (plotted for FBs, LBs, and groundwater samples) and UCL calculations (performed

for LBs) are nonparametric statistics that do not use the magnitudes of individual values but nevertheless require proper ranking of values in order from low to high. For S2437 pesticides, the proper ranking of results is complicated by (1) the general practice of reporting nondetections for environmental samples and FBs at the RL while reporting concentrations for confirmed detections below the RL and (2), for several pesticides, common reporting of nondetections for FBs and environmental samples at RRLs.

Regarding (1), because the NWQL reports all detections that meet identification criteria and relevant QC criteria at concentrations down to the lowest reportable value (table 1), which is typically 5 percent of the RL, while generally reporting nondetections for environmental samples at the RL, ranking values on the reported concentration alone (without

taking the remark code “<” into account) would result in nondetections incorrectly being assigned a higher rank than detections with concentrations below the RL. Therefore, all nondetections reported for environmental samples at the standard RL in effect at the time of analysis (even if that standard RL changed through time) were retained as nondetections with an assigned value—and a corresponding rank—below the lowest detection for that pesticide in 2013–18 for the purposes of calculating EDFs and UCLs. For LBs, the NWQL generally reports nondetections at zero, which was retained as the assigned value.

Regarding (2), reporting of nondetections at RRLs is generally uncommon for LBs, occurring for fewer than 10 percent of nondetections for 217 of 225 pesticides (maximum 15.5 percent for metconazole). However, reporting of nondetections at RRLs is common for several pesticides for FBs and groundwater samples, with 97 and 90 pesticides, respectively, having RRLs assigned for more than 10 percent of nondetections (47 and 49 pesticides, respectively, having RRLs assigned for more than 20 percent of nondetections). Similar to Shoda and others (2018; calibration standards listed in table 1 therein), for the current study, results reported with an RRL at or below the first calibration standard above the maximum standard RL for the study period were judged to provide a suitable level of resolution with a minimal increase in the potential for false-negative results and generally were included in data analysis (at values assigned as described in the following paragraph). All results reported with an RRL greater than the first calibration standard above the maximum standard RL for the study period were excluded from analysis, largely because of the greater loss of resolution and associated increase in the potential for false-negative results. Also similar to Shoda and others (2018), RRL results with no indication of the reason for the RRL were excluded from analysis. However, unlike Shoda and others (2018), RRL results with an indication of contamination in an associated LB, indicating uncertainty in quantification, were retained as nondetections and assigned a value at the RRL (this RRL reason is applicable to results reported for FBs and groundwater samples, but not for LBs). For LBs, only 10 pesticides had more than 5 percent of nondetections excluded under these rules, and only 1 pesticide had more than 10 percent of nondetections excluded (10.5 percent were excluded for 1H-1,2,4-triazole). For FBs and groundwater, 64 and 61 pesticides, respectively, had more than 5 percent of nondetections excluded under these rules, and 42 pesticides for each sample type had more than 10 percent of nondetections excluded. The general effect of exclusion of these nondetections is for some pesticides to potentially have a somewhat higher calculated detection frequency and somewhat higher concentrations calculated for some percentiles than would have been the case if these exclusions of nondetections (reducing the overall number of results while not affecting the number of detections) had not been made.

For the purposes of calculating EDFs and UCLs, the value—and corresponding rank—assigned for retained nondetections reported with an RRL for all sample types

depended on the reason provided by the NWQL in the comment and (or) VQC field for each particular RRL result. Shoda and others (2018) defined two broad categories for RRLs, a first category in which there is uncertainty in the identification of the pesticide and a second category in which there is uncertainty in the quantification of the pesticide. The first category can be subdivided into cases where the laboratory analyst observed no evidence that the pesticide was likely present in the sample and cases where some evidence indicated that the pesticide was likely present in the sample, but not all criteria for qualitative identification were met. For the current study, RRL results with a comment that specified interference or a change in instrument sensitivity during the run with no indication of evidence that the pesticide was present in the sample (accounting for at least 90 percent of all RRL comments for each sample type and typically accompanied by an RRL equivalent to the concentration of a calibration standard) were retained as nondetections with an assigned value—and a corresponding rank—below the lowest detection for that pesticide (or an assigned value of zero for LBs), equivalent to the value assigned for nondetections reported at the standard RL. RRL results with a comment indicating a failure to meet all acceptance criteria for qualitative identification (typically assigned an RRL equivalent to the concentration that would have been reported for the pesticide if its presence had been confirmed) were interpreted as indicating the possible presence of the pesticide in the sample at a concentration up to the RRL. These results generally were retained as nondetections assigned a value at the RRL and were allowed to be ranked accordingly for the nonparametric statistical procedure being used.

As discussed earlier in this section, UCLs were calculated only for LBs. For these calculations, one extra step was taken with RRL results after assigning their values as described in the preceding paragraph. Of nearly 406,700 LB results remaining after all previous data preparation, only 25 were RRL results that could not confidently be assigned the standard value of zero generally assigned to nondetections for LBs for ranking purposes. These 25 RRL results were dropped from the dataset for the purposes of determining UCLs because inclusion of results with more than 1 censoring level would require the use of a more complicated method than the binomial probability method for determining UCLs for the affected pesticides, and because excluding these few RRL results (affecting only 10 pesticides, and totaling fewer than 1 percent of results for any individual pesticide) was determined to have negligible effects on UCL calculations.

Data for Third-Party Blind Blank Samples

Data for BBSs were evaluated for contamination bias for individual pesticides by examining the frequency and concentration of false-positive results. BBS results were compared with the data quality objective that the false-positive occurrence rate be no higher than 1 percent when considering only false-positive detections reported as concentrations

greater than the maximum DL. This is a useful threshold for assessing method performance because the DL is defined as the lowest concentration with a false-positive risk of no more than 1 percent, and false-positive risk is expected to be higher at concentrations below this. After compilation of the dataset, in which any detection indicated a false-positive result, no additional data preparation was required to perform this analysis.

Data for Laboratory Reagent Spikes, Field Matrix Spikes, and Third-Party Blind Spikes

Data for LRSs, FMSs, and BSSs were evaluated for bias and (or) variability for individual pesticides by using statistical summaries and graphical analysis of spike recoveries. Specifically, tables of summary statistics that were created for LRS, FMS, and BSS recoveries included minimum, maximum, median, and mean recoveries, along with the standard deviation (SD), relative standard deviation (RSD; the ratio, in percent, of SD to mean concentration), and selected additional percentiles of recoveries for each pesticide. The 75th and 25th percentiles were used to calculate the F-pseudostigma, a nonparametric measure of variability (less sensitive to outliers than the RSD) equal to the difference in these percentiles divided by 1.349 (Hoaglin and others, 1983). With respect to bias, recovery results for all spike types were compared with the data quality objective that median recovery fall within the range of 70–130 percent, with 100 percent being perfect recovery; this recovery range matches the initial acceptance criteria set by Sandstrom and others (2015) for the analytical method and was judged acceptable for NAWQA objectives for the pesticide data for groundwater. With respect to variability, recovery results for LRSs and BSSs, which each provide numerous measurements for the same matrix and the same or similar spike concentrations, were compared with the data quality objective that F-pseudostigma, commonly used by the NWQL to assess variability (USGS, 2005, 2007), was equal to no more than 30 percent; alternatively, RSD can be used in place of F-pseudostigma.

LRS, FMS, and BSS recoveries were each plotted for individual pesticides by analysis date to illustrate the range of typical recoveries and changes in recoveries through time. Kernel regression smoothing curves (Altman, 1992) were included on the graphs to illustrate broad changes in recovery. These curves were generated with the *lokern* package (Herrmann, 2016) in R software (R Core Team, 2019) by using an assigned bandwidth of 180 days, which means that recovery values within a 6-month window influence the smooth at each value. Figure 4 shows graphs of LRS, FMS, and BSS recoveries for imazamox, aldicarb sulfoxide, and acetochlor sulfonic acid during the 2013–18 time period; graphs for other pesticides are available in Bexfield and others (2020, PlotGroup3). Box plots also were generated in R for individual pesticides to directly compare recoveries for LRSs and FMSs and evaluate differences in results for the two

types of spikes that could reflect matrix effects or degradation of a compound between spiking and analysis; figure 5 shows box plots for naled and orthosulfamuron (graphs for other pesticides are available in Bexfield and others, 2020, PlotGroup4). Results for BSSs were not included in this direct comparison because the spike solutions (the individual mix of compounds and the concentrations) used for BSSs typically are not the same as those used for LRSs and FMSs. BSSs typically are spiked with compounds at varying concentrations near the RL, and below the nominal concentration of 250 ng/L for LRSs, but a small number of BSSs are spiked at concentrations several times higher (varying by compound in number and concentration); for about 25 percent of compounds, the maximum concentration is at least 10 percent of the HHB. Although the generally small number of BSSs with these higher concentrations does not allow for a robust analysis of any potential relation between concentration and recovery, graphs that show recovery plotted against spiked concentration are available on the QSB Organic Blind Sample Project website (<https://qsb.usgs.gov/OBSP/index.html>).

For LRSs, recoveries were retrieved directly from the NWQL Laboratory Information Management System, where they had been determined as the concentration measured in each LRS expressed as a percentage of the known concentration added to the sample. The primary data preparation step required for the compiled LRS dataset was to remove results for LRSs that had been spiked with NWQL spike lot 91219, which was used in 2013–14 and was determined by Shoda and others (2018) to have unusually low recoveries and high variability for certain analytical groups relative to other spike lots, possibly as a result of degradation of some pesticides soon after lot certification or a preparation or handling error.

For BSSs, recoveries were retrieved directly from the QSB database, where they had been determined as the concentration measured in each BSS expressed as a percentage of the known concentration added to the sample. BSSs were additionally evaluated for bias by calculating occurrences of false-positive and false-negative results, which had been identified as such directly in the QSB database. As mentioned in the preceding section, the objective for false-positive occurrence was a rate no higher than 1 percent at concentrations greater than the maximum DL. For false-negative occurrence, the data quality objective was a rate no higher than 10 percent for samples that were spiked at a concentration above the maximum RL. Results with a VQC indicating contamination in associated LBs were excluded from analysis of recoveries because of the potentially large effect that the associated uncertainty in quantification of the concentration could have on the recovery determination. Data were evaluated for pesticides that were spiked into at least 10 BSSs. Because BSS results had been reviewed by QSB personnel for any potential spike solution issues, no further data preparation was performed prior to analysis.

Using the equations provided in Mueller and others (2015), FMS recoveries were calculated on the basis of concentrations

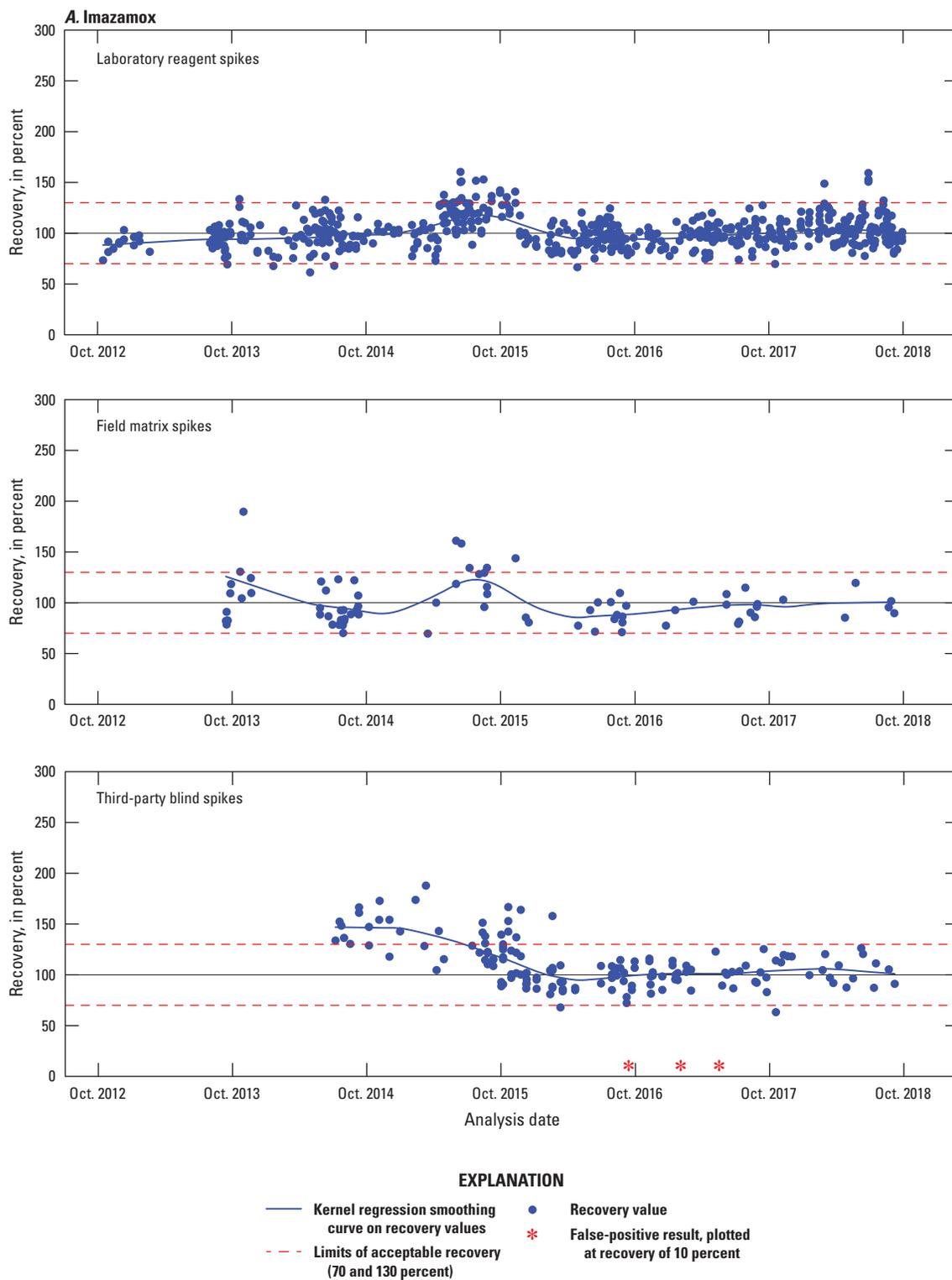


Figure 4. Recovery values in laboratory reagent spikes, field matrix spikes, and third-party blind spikes plotted by analysis date for *A*, imazamox; *B*, aldicarb sulfoxide; and *C*, acetochlor sulfonic acid.

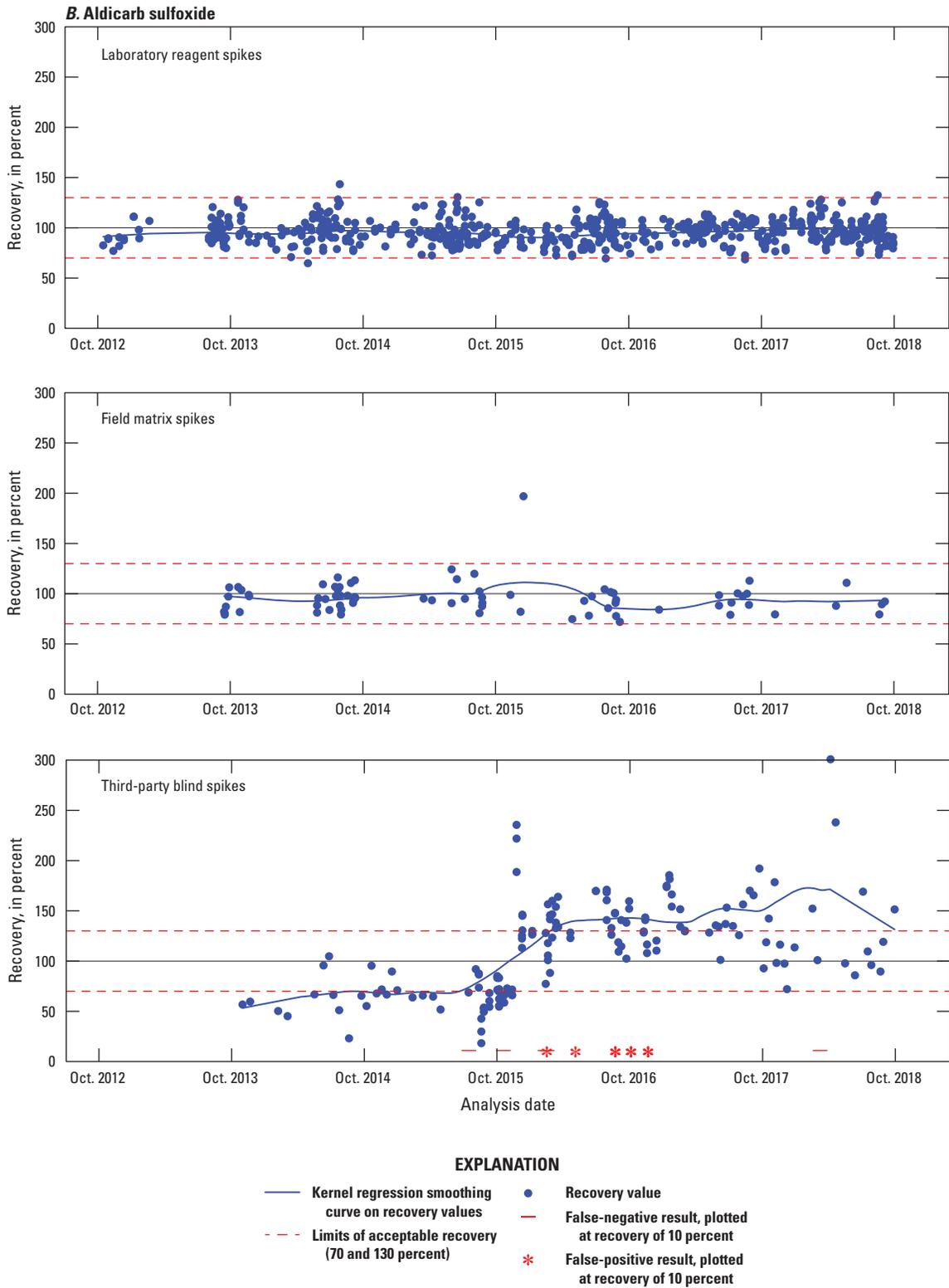


Figure 4. Recovery values in laboratory reagent spikes, field matrix spikes, and third-party blind spikes plotted by analysis date for *A*, imazamox; *B*, aldicarb sulfoxide; and *C*, acetochlor sulfonic acid.—Continued

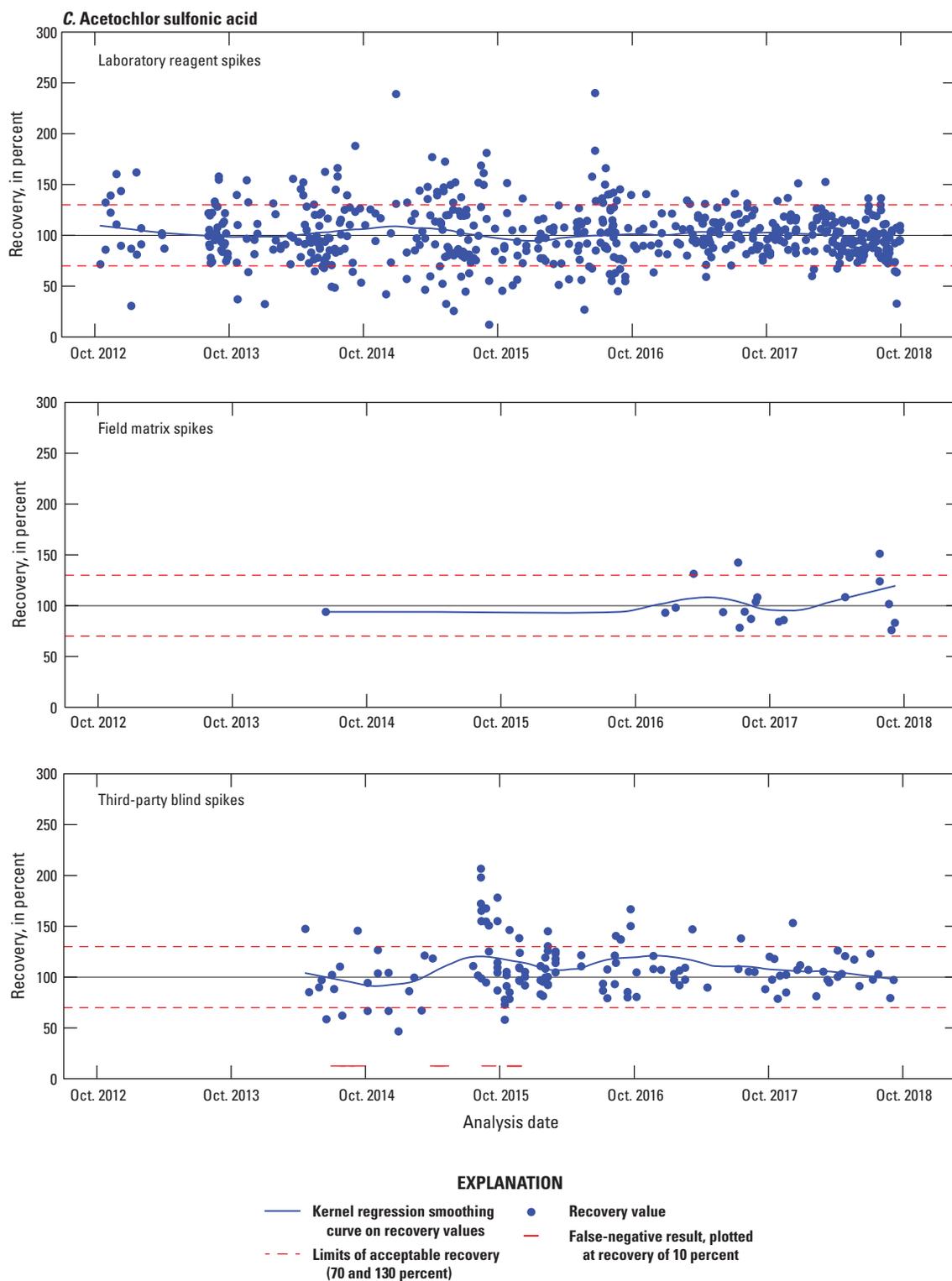


Figure 4. Recovery values in laboratory reagent spikes, field matrix spikes, and third-party blind spikes plotted by analysis date for *A*, imazamox; *B*, aldicarb sulfoxide; and *C*, acetochlor sulfonic acid.—Continued

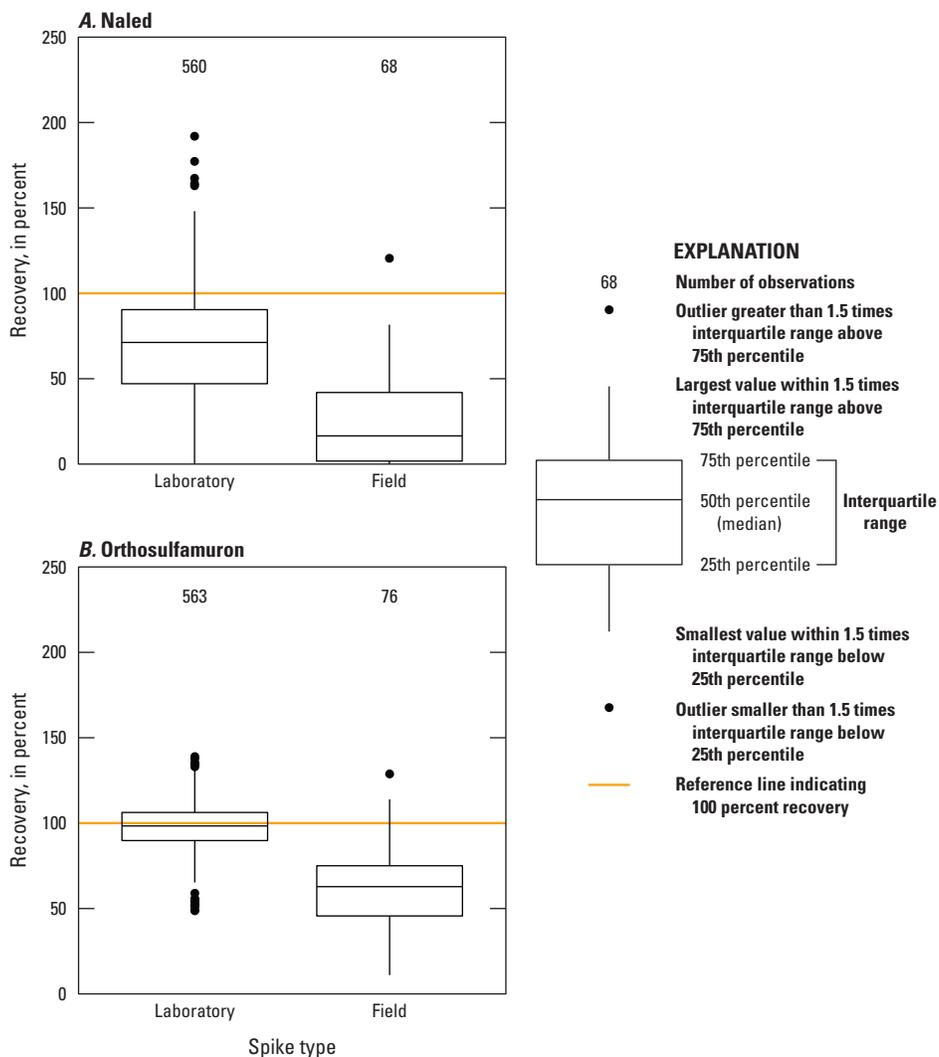


Figure 5. Comparison of recovery values in laboratory reagent spikes and field matrix spikes for *A*, naled and *B*, orthosulfamuron.

measured in the FMSs and their paired groundwater samples that were not spiked. Specifically, recovery was determined as the difference of the concentration of the pesticide in the FMS and the concentration in the paired groundwater sample (assumed to be zero if the result was a nondetection) divided by the expected concentration and then multiplied by 100 to express the result as a percent. The expected concentration was calculated as the volume of spike solution added to the sample (assumed to be 100 μL , as directed by protocols) multiplied by the concentration of the pesticide in the spike solution (obtained from the NWQL certificate corresponding with the lot specified in NWIS) and divided by the volume of the spike sample (reported by the NWQL).

Data preparation for FMS recovery calculation mostly followed the approach described by Shoda and others (2018). Specifically, results reported for either the FMS or the paired groundwater sample with an RRL greater than the first calibration standard above the maximum standard RL were excluded from analysis. RRL results with no indication of the reason for the RRL also were excluded. In addition, all results

(not just RRL results, as targeted by Shoda and others, 2018) where the LMS and (or) environmental sample had an indication of contamination in associated LBs were excluded from analysis because of the potentially large effect that the associated uncertainty in quantification of the concentration could have on the recovery calculation. Results were removed for FMSs that had been spiked with the likely degraded spike lot 91219 and those that had been spiked with any lot that had expired prior to sample collection, indicating a high potential for degradation. Results also were removed when the expected spike concentration was lower than the maximum RL listed in table 1 for that pesticide, indicating that the method might not be capable of detecting the spike concentration, or when the concentration in the paired groundwater sample exceeded the expected spike concentration (that is, when the background concentration was high), resulting in increased uncertainty in recovery analysis. If the recovery calculated for a pesticide in an FMS was negative, which is not technically possible and reflects a smaller reported concentration in the FMS than in the paired groundwater sample typically due to high

analytical variability and (or) concentrations below the RL, the recovery value was set to zero prior to graphing or calculating summary statistics for results. Five samples were removed prior to further data analysis because of evidence of a potential spiking error, which included an overall lack of detections and (or) laboratory or field comments indicating that the sample was not in fact a spike, recoveries of about two or more times typical values indicating that the sample was likely double-spiked, or a pattern of detections and concentrations indicating that a spike solution intended for a different pesticide schedule likely was mistakenly used.

Data for Field Replicates

Data for FRs were evaluated for variability for individual pesticides by calculating measures of consistency in detections and differences in concentrations within replicate pairs in accordance with procedures described by Mueller and others (2015). To ensure that these results were reasonably representative of variability, calculations of consistency in detection were made only for pesticides that had 10 or more replicate pairs with a detection in at least 1 sample of the pair, and calculations of differences in concentration were made only for pesticides that had 10 or more replicate pairs with detections in both samples of the pair. The mean detection rate for all replicate pairs having at least one detection (where the detection rate for individual pairs with detections in both samples is 100 percent and for individual pairs with a detection in only one sample is 50 percent) was calculated to evaluate variability in analyte detection (Martin, 2002; Mueller and others, 2015). In addition, the percentage of replicate pairs with inconsistent detections was calculated as the number of replicate pairs with inconsistent detections (that is, one detection and one nondetection) divided by the total number of replicate pairs minus the number of pairs with consistent nondetections (Martin, 2002; Mueller and others, 2015); results using the total number of replicate pairs also are provided in the “Variability Indicated by Results for Field Replicates” section of this report. A piecewise-linear model used by Mueller and Titus (2005) and described by Mueller and others (2015) was used to estimate variability in concentration. This approach divides concentrations into two ranges. In the low concentration range, the SD of replicates generally is uniform, and the average SD is used to estimate variability. In the high concentration range, the RSD generally is uniform, and the average RSD is used to estimate variability. An appropriate boundary between the low and high concentration ranges is selected using graphs of SD and RSD against mean concentration (Mueller and others, 2015). In some cases, either SD or RSD is fairly uniform throughout the range of available concentrations and is used as the sole measure of variability.

The evaluation of FRs was intended to characterize variability in results as reported by the NWQL for samples that were designed to be essentially identical. Therefore, all RRL values were retained as reported, and no data processing was needed after compilation (that is, there was no need to rank data or to treat nondetections with different reporting levels differently).

Results of Data-Quality Assessment

Results of the data-quality assessment for pesticide results determined by S2437 indicate that most pesticides have acceptable bias and variability with respect to data quality objectives specified in the “Data Preparation and Evaluation” section of this report. However, for some pesticides, bias and (or) variability are high enough that potential effects on the interpretation of groundwater data need to be considered; this section focuses on those results.

Outcome of Laboratory Verification

As discussed in the “Data Preparation and Evaluation” section of this report, laboratory verification was requested for groundwater detections occurring in temporal and geographic patterns that were judged to be unlikely to occur under environmental conditions. Of the 107 detections targeted for review, the NWQL determined that 22 should have been reported as nondetections because they did not meet all qualitative criteria and that 57 would have been reported as nondetections under the interference censoring rules that went into effect on May 21, 2015 (as described in the “Laboratory Analysis by Schedule 2437” section of this report). In addition, the NWQL determined that 13 remaining detections and 2 of the values changed to nondetections after reevaluation should have included a VQC of “i” to indicate interference in associated LBs. For the purposes of this study, the reevaluated results were determined to be more likely to represent true groundwater conditions than the results that were originally reported; therefore, the reevaluated results are used in this report and listed in table 1 of Bexfield and others (2020). The 92 results that had a change in value and (or) VQC are flagged in that table, allowing the affected compounds (5 parents and 8 degradates) and range of analysis dates (mostly periods of a few days between July and October of 2013–15) to be easily identified. The months of analysis of reevaluated results are consistent with the typically increased frequency of sample collection between May and September of each year. The fact that all detections selected for laboratory verification were reported for samples analyzed between July 2013 and April 2015 implies that the interference censoring rules implemented in May 2015 and (or) general method performance improvements in the first 2 years of S2437 implementation may have reduced the potential for reporting of false-positive results for at least some pesticides after about April 2015.

Contamination Bias

FB and LB results and corresponding implications for contamination bias are presented in this section of the report. Results also are provided and discussed for BBSs submitted by the QSB to the NWQL for analysis during 2018 and for 2013–18 BSSs.

Field Blanks

Results for FBs indicate that there is little potential for groundwater results to be substantially affected by contamination from various aspects of sample collection, handling, and analysis. After data preparation, the number of FB results available for an individual pesticide ranges from 85 to 220 (220 FBs equate to about 11 percent of groundwater samples collected for NAWQA in 2013–18) (table 3); numbers of results substantially less than 220 were a consequence of the data preparation step that excluded RRL results reported at values greater than the first calibration standard above the maximum standard RL. Of 225 pesticides, 192 were not detected in any FBs (table 3); 23 pesticides were each detected only once, equivalent to a detection frequency of about 0.5 percent. Eight pesticides were each detected in two FBs, equivalent to a detection frequency of about 0.9–1.2 percent depending on the pesticide and its associated number of results. The largest numbers of detections were reported for desulfinylfipronil, which was detected in 3 of 217 FBs (1.4 percent; detections in FBs submitted by 3 different field crews), and metolachlor, which was detected in 4 of 217 FBs (1.8 percent; detections in FBs submitted by 4 different field crews); both of these pesticides were also detected in at least 1 groundwater sample (table 3). Given these low detection frequencies in FBs, the potential for contamination bias to be traced to different sources, such as from the use of portable versus dedicated pumps for groundwater sampling or from collection of samples in urban compared with agricultural areas, was not investigated.

Most of the detected pesticides had maximum FB concentrations below their laboratory limits (DL or RL) and several orders of magnitude lower than their corresponding HHBs (table 3). Nine pesticides were detected at least once at a concentration above their maximum DLs; all three desulfinylfipronil detections were above the maximum DL. Three pesticides (pyrimidinol, cis-permethrin, and trans-permethrin) were detected at least once at a concentration above their maximum RLs; the maximum concentration reported for pyrimidinol was 35.8 ng/L, which was the highest FB concentration reported for any pesticide (most maximum concentrations were lower than 10 ng/L). For all but 3 of the 22 pesticides with FB detections and an available HHB, the maximum FB concentration was at least 3 orders of magnitude lower than the HHB. The largest ratios of maximum FB concentration to HHB (about 0.0022–0.0025) were for desulfinylfipronil, cis-permethrin, and trans-permethrin, indicating that the observed levels of contamination bias should not substantially affect the suitability of reported groundwater concentrations of any pesticide for comparison with its HHB or 10 percent of its HHB.

As mentioned in the “Data Preparation and Evaluation” section of this report, 90-percent UCLs for the 90th and 95th percentiles of FB concentration were not calculated. Given the low detection frequency for all pesticides in FBs (1.8 percent or less, even after excluding some nondetections for several

pesticides because of high RRLs), the sample concentrations corresponding to the ranks for these UCLs would have been nondetections.

Laboratory Blanks

Results for LBs indicate the potential for groundwater results for some pesticides to be substantially affected by contamination from laboratory processes. After data preparation, the number of LB results available for an individual pesticide ranges from 1,470 to 1,886 (table 3); numbers of results substantially less than 1,886 were a consequence of differences in the number of LBs analyzed for individual pesticides and (or) the data preparation step that excluded RRL results reported at values greater than the first calibration standard above the maximum standard RL. Of 225 pesticides, 19 were not detected in any LBs. A total of 125 pesticides were each detected fewer than 10 times, or in less than 0.5 percent of LBs; 49 pesticides were each detected in 0.5 to less than 3 percent of LBs. The remaining 32 pesticides were each detected in 3–91 percent of LBs. The nine pesticides detected in more than 10 percent of LBs each were pyridaben (12.0 percent), metalaxyl (13.4 percent), piperonyl butoxide (19.2 percent), nicosulfuron (19.9 percent), atrazine (21.0 percent), etoxazole (24.2 percent), pyrimidinol (29.1 percent), cis-permethrin (31.0 percent), and fenbutatin oxide (91.5 percent). All of these pesticides except etoxazole were detected in at least one groundwater sample (table 3).

More than half of the detected pesticides had maximum LB concentrations higher than one or both of their laboratory limits, but most had maximum LB concentrations several orders of magnitude lower than their corresponding HHBs (table 3). Of the 206 pesticides detected in LBs, 120 had a maximum concentration above the maximum DL, and 74 had a maximum concentration above the maximum RL. The maximum LB concentration exceeded 10 ng/L for 70 pesticides and exceeded 100 ng/L for 10 pesticides (ranging up to 513 ng/L for 3-phenoxybenzoic acid, for which no detections were reported in groundwater samples). Of the 112 pesticides with LB detections and an available HHB, the maximum LB concentration was at least 3 orders of magnitude lower than the HHB for 75 pesticides and was at least 2 orders of magnitude lower than the HHB for 22 additional pesticides. Therefore, for all but 15 pesticides, even the maximum observed laboratory contamination bias would not substantially affect the suitability of reported groundwater concentrations of a pesticide for comparison with its HHB or 10 percent of its HHB. Fourteen of the 15 highest ratios of maximum LB concentration to HHB range from 0.012 to 0.070, for dicotophos (0.012), fipronil (0.012), diuron (0.014), diazinon (0.014), chlorpyrifos oxon (0.016), terbufos (0.019), profenofos (0.021), tebupirimfos (0.022), atrazine (0.022), desulfinylfipronil (0.023), bromoxynil (0.035), propargite (0.036), oxyfluorfen (0.065), and cyanazine (0.070). These ratios imply the potential for the maximum observed laboratory contamination bias (that is, bias of

about 1–7 percent of the HHB) to affect the suitability of reported groundwater concentrations of these 12 pesticides for comparison with 10 percent or less of their corresponding HHB but should not affect their suitability for identifying exceedances of the HHB; in addition, terbufos, bromoxynil, propargite, and cyanazine were not detected in any groundwater samples. The ratio of maximum LB concentration to HHB was 1.20 for fentin, and therefore, if the maximum observed LB concentration of 24 ng/L for fentin contributed to the reported concentration of this pesticide in groundwater, this contamination bias would make the groundwater result unsuitable for comparison to its HHB of 20 ng/L. However, it is worth noting for fentin that the maximum DL is 75 percent of the HHB and the maximum RL is 150 percent of the HHB, meaning that this pesticide is difficult to identify and quantify by this method at concentrations of potential relevance to human health.

It should be noted that detection frequencies and maximum reported concentrations for individual pesticides on S2437 were substantially lower in FBs than in LBs, even though FBs necessarily are subjected to the same laboratory processes that can affect LBs and it is generally accepted that the bias of field data is unlikely to be less than the bias of the analytical method (Mueller and others, 2015). Although these differences have not been directly investigated for the datasets used in this study, it seems likely that they result from a combination of three factors: (1) a larger population of LBs compared with FBs, which increases the chances for the maximum potential contamination to be observed; (2) the practice of reporting of LB detections at concentrations below the lowest reportable value applicable to FBs (table 1); and (3) the laboratory practice of set-by-set censoring described in the “Laboratory Analysis by Schedule 2437” section of this report (as mentioned in that section, this type of censoring likely occurs more often for S2437 than for previous NWQL pesticide schedules because instrument blanks are taken into consideration). As discussed by Medalie and others (2019), the detection frequency of some pesticides in LBs might be greater than the detection frequency observed for FBs and environmental results in the same set as a consequence of set-by-set censoring because LBs are the only type of sample for which pesticide detections are not censored by the laboratory. Data users typically are not aware when a raw FB result was a detection that was reported and (or) published as a nondetection because of set-by-set censoring, unless the concentration was more than three times the concentration in the associated LBs and a VQC of “v” was added to the FB result. Given that FBs and groundwater samples are treated the same for laboratory censoring purposes, it is likely that the practice of set-by-set censoring also results in contamination bias from laboratory processes being less frequent and lower in magnitude for reported groundwater concentrations than LB results would imply in the absence of this censoring practice; however, the precise effects on groundwater values cannot be assessed without knowing the raw measurements prior to censoring. Whereas set-by-set censoring has been

shown to effectively address the effects of deterministic and semideterministic contamination (that is, contamination observed in LBs that affects all or some of the 15 associated environmental samples in a sample set), it does not necessarily address episodic contamination (characterized by clusters of detections in time) or random contamination, which were shown to be more common types of contamination for certain pesticides analyzed using previous NWQL pesticide methods that utilize gas chromatography and mass spectrometry (Medalie and others, 2019).

Further examination of LB results by analysis date indicated that patterns of detection can differ substantially over time, with differences by water year being quite common even for pesticides with relatively high detection frequencies in LBs (for example, see atrazine and metolachlor in fig. 2A and B). Because such patterns appear to be quite common (as supported by graphs in PlotGroup1 of Bexfield and others, 2020), further analysis is presented by water year. Factors resulting in large variability in detection frequencies by water year for multiple pesticides have not been identified. Cases where short episodes of contamination in LBs (over a period of up to 3 months or so) closely overlap with unusual clusters of detections in groundwater are rare. One example of this pattern of potential short-term contamination of groundwater samples by laboratory processes occurred for diketonitrile isoxaflutole in early September 2014 (fig. 2C); a review by the NWQL found that most of the detections for this pesticide in groundwater samples analyzed at this time met reporting criteria.

As described in the “Data Preparation and Evaluation” section of this report, for pesticides that were detected in LBs, 90-percent UCLs for the 90th, 95th, and 99th percentiles of LB concentration were calculated (after dropping 25 RRL results that could not confidently be assigned the standard value of zero for nondetections in LBs). Because examination of LB detections by analysis date indicated that detection frequency and concentration could vary substantially by water year, the UCLs were calculated for individual water years. Table 4 (available for download at <https://doi.org/10.3133/sir20205072>) lists the results for pesticide/water year pairs that had a quantifiable 90-percent UCL for the 90th, 95th, and (or) 99th percentile; all other results were zero, indicating that a nondetection corresponded with the rank of the UCL. The EDFs for atrazine in 2014 and fipronil sulfone in 2015 (fig. 3) compare distributions of concentrations reported for FBs, LBs, and groundwater samples at high percentiles. EDFs for other pesticides and water years that had a quantifiable 90-percent UCL for the 99th percentile of LBs and at least one groundwater detection are available in Bexfield and others (2020, PlotGroup2). Note that the percentiles at which concentrations plot on the EDFs will not exactly match percentile values listed in table 4 because values in the table correspond with the UCL for each percentile. In this section, the magnitudes of the 90-percent UCLs for the 90th and 95th percentiles are discussed in relation to laboratory limits and HHBs; the UCL for the 99th percentile, which has been

discussed by Fram and Stork (2019) as one potential threshold for data censoring and would constitute the most conservative percentile for this purpose, is not specifically discussed here, but all information needed to perform the same comparisons is provided in table 4.

Thirty-five pesticides have a quantifiable 90-percent UCL for the 90th percentile for at least 1 water year, and 62 pesticides have a quantifiable 90-percent UCL for the 95th percentile for at least 1 water year (table 4). Table 4 also lists the minimum concentration reported for a groundwater sample for each pesticide during each water year. The 90-percent UCL for the 90th percentile equals or exceeds the minimum reported groundwater concentration in at least 1 water year for 12 pesticides: atrazine, cis-permethrin, diuron, fenbutatin oxide, fluometuron, linuron, metalaxyl, nicosulfuron, pendimethalin, piperonyl butoxide, pyridaben, and tebuthiuron. The 90-percent UCL for the 95th percentile equals or exceeds the minimum reported groundwater concentration in at least 1 water year for 28 pesticides, consisting of those 12 plus these 16 additional pesticides: 2,4-D, azoxystrobin, bromacil, carbaryl, diazinon, diflufenzuron, fipronil sulfide, fipronil sulfone, imazethapyr, methoxyfenozide, metolachlor, norflurazon, prometon, propoxur, siduron, and sulfosulfuron. Therefore, at least one concentration reported for these pesticides during the water years noted in table 4—and possibly several reported concentrations, particularly for more frequently detected pesticides—could result from laboratory contamination bias, rather than being representative of groundwater conditions, indicating the need to consider data censoring for certain study objectives. The 90-percent UCL for the 90th percentile exceeds the maximum DL during 1 water year each for 2,4-D, ametryn, and propiconazole and during 2 water years each for atrazine and fenbutatin oxide; the 90-percent UCL for the 90th percentile exceeds the maximum RL during 2013 for atrazine. The 90-percent UCL for the 95th percentile exceeds the maximum DL during at least 1 water year for 17 pesticides: 2,4-D, 2-hydroxyatrazine, ametryn, atrazine, cis-permethrin, diazinon, diuron, fenbutatin oxide, fipronil sulfide, imazaquin, metalaxyl, metconazole, pendimethalin, propiconazole, propoxur, pyriproxyfen, and trans-permethrin. The 90-percent UCL for the 95th percentile exceeds the maximum RL during 1 water year each for ametryn, atrazine, and diazinon.

Of the 33 pesticides with a 90-percent UCL for the 90th percentile for at least 1 water year and an available HHB, the 90-percent UCL for the 90th percentile was at least 2 orders of magnitude lower than the HHB for all pesticides and water years. Similarly, of the 54 pesticides with a 90-percent UCL for the 95th percentile for at least 1 water year and an available HHB, the 90-percent UCL for the 95th percentile was at least 2 orders of magnitude lower than the HHB for all pesticides and water years. Therefore, laboratory contamination bias in up to at least 95 percent of groundwater samples, with 90-percent confidence, should not substantially affect the suitability of reported groundwater concentrations for comparison with the HHB of any pesticide or 10 percent of the HHB for any pesticide.

The EDF plot of reported groundwater concentrations relative to FB and LB concentrations for atrazine in 2014 (fig. 3A) illustrates a case when the distribution of concentrations measured in groundwater differs substantially from the distributions of concentrations measured in blanks. In particular, for any given percentile, the concentration in groundwater is substantially higher than the concentration in blanks for that same pesticide and water year, indicating that the effects of contamination bias on reported groundwater concentrations likely are minor. Plots provided in PlotGroup2 of Bexfield and others (2020) support the occurrence of similar differences in distributions for most pesticides and water years with detections reported for groundwater and LBs. However, for some pesticides and water years, such as for fipronil sulfone in 2015 (fig. 3B), the distributions plot quite close or overlap, with the concentration in LBs at some percentiles being higher than the concentration in groundwater. When these patterns are observed, examination of the precise timing of groundwater detections relative to blank detections, along with consideration of censoring of some groundwater detections, is warranted to ensure that groundwater sample results do not have an unacceptably high probability of being substantially affected by contamination bias.

Third-Party Blind Blank Samples and Blind Spike Samples

Available results for BBSs, which were submitted by the QSB to the NWQL for analysis during 2018, totaled 15–20 results per pesticide and indicated no clear contamination bias for all but 1 pesticide. Only propoxur had any false-positive results. The 2 propoxur detections (reported at 1.9 and 26.9 ng/L) out of 18 BBSs give a false-positive occurrence of about 11 percent, which is substantially higher than the data-quality objective of 1 percent; however, the reported concentration for 1 detection was lower than the maximum DL, meaning that the false-positive occurrence at concentrations above the maximum DL is 5.6 percent. The two propoxur detections occurred on March 19, 2018, and April 2, 2018, indicating potential contamination bias during this period. However, internal NWQL quality-assurance testing of the 20-mL vials used for collection of samples for analysis by S2437 indicated that propoxur was present at low concentrations in the sampling vials used in 2016 and 2018, which could have caused at least some of the reported false-positive results in 2018 BBSs and in 2016–18 BSSs (discussed in the following paragraph).

Available results for 223 pesticides that were not spiked into individual BSSs submitted for analysis during 2013–18, numbering from 24 to 271 results per pesticide, indicate that, when considering only false-positive detections reported at concentrations greater than the maximum DL, false-positive occurrence exceeds 1 percent for 31 pesticides (including propoxur at 2.0 percent), 29 of which were detected in at least 1 groundwater sample (table 5, available for download at <https://doi.org/10.3133/sir20205072>). (If all false-positive detections were included, false-positive occurrence would

exceed 1 percent for 50 pesticides.) These results indicate the potential for positive bias to affect detections reported for these 29 pesticides in groundwater samples analyzed during the same time period. The reason for a higher false-positive occurrence for these pesticides based on BSSs relative to BBSs could not be definitively determined, although possible explanations include the relatively low number and restricted time period of BBSs, interferences or misidentification of pesticides in BSSs, and (or) issues with pesticides being present in BSSs when they were not intended to be included in spike solutions.

In 2018, in contrast to BBSs, BSSs indicated that 16 pesticides (including propoxur) had at least 1 false-positive result and that 5 pesticides (acetolachlor, chlorpyrifos, oxamyl, phthalazinone, and pyridaben) had at least 2 false-positive results. Given the matching time period and the similar number of BBS samples (16–18) and BSS samples that were not spiked (9–27) for the 16 pesticides with false-positive results for BSSs in 2018, it appears that factors other than the time period and number of BBS samples explain most differences in false-positive results between sample types. Note that, although combining 2018 BBS and 2013–18 BSS results would reduce the false-positive occurrence (considering only concentrations greater than the maximum DL) for seven pesticides from just above to just below 1 percent, combining these results could skew the results toward laboratory performance in 2018.

For 2013–18 BSS results, 10 of the 31 compounds with false-positive occurrence at concentrations greater than the maximum DL exceeding 1 percent had VQCs and (or) comments noting the presence of interference for 1 or more false-positive results, indicating that interference might affect some results. The compounds with the two highest false-positive occurrences at concentrations above the maximum DL are both degradates (table 5), which could indicate that at least some of the higher false-positive occurrence rates result partially or mostly from degradation of parent compounds spiked into BSSs, rather than resulting primarily or entirely from laboratory method performance. In addition, although 15 of the 31 compounds with high false-positive occurrence were detected in at least 1 percent of LBs (table 3), providing additional evidence of the potential for high bias to affect environmental samples, the other 16 compounds (including 10 degradates, 7 of which generally were accompanied by a known parent compound in BSSs with false-positive results) were detected in fewer than 1 percent of LBs and FBs; also, all but 1 of these compounds (simazine) were detected in fewer than 1 percent of groundwater samples (table 3). Along with a lack of indicators of interference for 15 of these 16 compounds, these results do not provide clear support for the BSS results and indicate that spiking issues or degradation after spiking might be contributing factors to false-positive occurrence for these compounds.

Focusing on the 17 pesticides that were detected in at least 1 groundwater sample and that have false-positive occurrences greater than 2 percent for BSSs when considering

only detections reported with concentrations greater than the maximum DL (table 5), most have no detections in groundwater samples occurring within a month of a false-positive result (Bexfield and others, 2020, tables 1 and 8, data plotted in PlotGroup1 and PlotGroup3). Exceptions are 2-hydroxyatrazine, acetochlor, atrazine, methomyl, propazine, and pyrimidinol, which each have one or more detections in groundwater samples occurring within a month of a false-positive result. These pesticides might be more likely than others to have one or more groundwater results affected by high bias. Four of these pesticides (2-hydroxyatrazine, atrazine, propazine, and pyrimidinol) were detected in at least 1 percent of LBs, which provides additional evidence of the potential for high bias to affect some environmental samples, but detections of acetochlor and methomyl in LBs were uncommon.

Bias and Variability Indicated by Results for Spike Samples

LRS, BSS, and FMS results are presented in this section of the report, along with corresponding implications for method performance with respect to bias and variability.

Laboratory Reagent Spikes

Recovery results for LRSs indicate that most pesticides on S2437 meet data quality objectives for method performance for both bias and variability. After data preparation, the number of LRS results available for an individual pesticide ranges from 389 to 557, with only 3 pesticides having fewer than 520 results (table 6, available for download at <https://doi.org/10.3133/sir20205072>). The median recovery for individual pesticides ranges from 70.1 to 108.4 percent (table 6), which is within the data quality objectives of 70–130 percent. Mean recoveries fall within a similar range, although the mean recovery for one pesticide, lactofen, is relatively low at 69.2 percent. Of the 119 pesticides with HHBs, 89 have a median recovery between 90 and 110 percent, implying that reported concentrations for these pesticides typically would have very little method bias and could accurately represent any exceedances of HHBs that may occur for groundwater, assuming similar method performance between matrices. The other 30 pesticides with HHBs have a median recovery below 90 percent (but above 70 percent), indicating a small to moderate low bias that could cause some reported concentrations to be categorized as being at or below the HHB when the actual concentration in groundwater is slightly greater than the HHB. All pesticides have at least one recovery value (minimum or maximum) outside the 70- to 130-percent range; it is not known whether these extreme values are representative of occasional issues with method performance or a sample error of some type. It should be noted that these overall results reflect method performance for reagent water and that bias and variability for an individual pesticide can differ for a typical groundwater matrix, as discussed in the “Field Matrix Spikes” section of this report.

For 223 of the 225 pesticides on S2437, the F-pseudosigma values of recovery met the data quality objective of being no more than 30 percent (table 6). F-pseudosigma values for the other two pesticides, fenbutatin oxide and naled, are 30.6 and 32.0 percent, respectively, implying relatively high variability in recovery and, therefore, potentially in reported concentrations; both of these pesticides have an HHB. These pesticides are both among the 19 pesticides that have routinely been reported since method implementation with a VQC of “m,” meaning that results determined for the compound by S2437 are highly variable (table 1). Using RSD as an alternative measure of variability, eight pesticides exceed 30 percent; these eight pesticides, which include fenbutatin oxide and naled, also have been known since method implementation to have highly variable results in some matrices.

The graphs in figure 4, as well as the graphs for additional pesticides in Bexfield and others (2020, PlotGroup3), illustrate changes in the typical LRS recovery for individual pesticides throughout 2013–18 (note that the gap in data points for part of 2013 is the consequence of having to drop results for LRSs that used spike lot 91219, as described in the “Data Preparation and Evaluation” section of this report). For some pesticides, time periods with a relatively high frequency of particularly low or high recoveries are evident (for example, see the relatively high recoveries for imazamox in late 2014 in fig. 4A). These same general patterns in method performance might be reflected in results reported for groundwater samples and could be an important consideration for certain uses of the groundwater data, such as for examining changes in concentration over time.

Third-Party Blind Spike Samples

Relative to results for LRSs, results for the 184 pesticides that were spiked into at least 10 BSSs submitted by the QSB for analysis by S2437 in 2013–18 indicate that more pesticides had bias and variability exceeding data quality objectives, potentially because of contributions from factors other than method performance. The number of BSS results available for an individual pesticide in the final dataset ranges from 20 to 205 (table 7, available for download at <https://doi.org/10.3133/sir20205072>). The median recovery for individual pesticides ranges from 31.3 to 348.9 percent (table 7), with 163 pesticides having a median recovery between 70 and 130 percent. The median recovery exceeds 130 percent for phorate sulfoxide (142.4 percent) and phthalazinone (348.9 percent). Given that the median recovery for these compounds does not exceed 100 percent in LRSs, it seems possible that the parent compounds (phorate and diflufenzopyr, respectively, both showing lower recovery for BSSs compared with LRSs) that are spiked into the same samples could be transforming into these degradate compounds between spiking and sample analysis. Degradation of parent compounds within BSSs might also contribute to the relatively high but still acceptable median recoveries (about 110 percent or greater)

for other degradate compounds. The median recovery is lower than 70 percent for 19 compounds, 13 of which are parent compounds. Five of these parent compounds (bifenthrin, lactofen, naled, and cis- and trans-permethrin) are among the seven that have median recoveries below 80 percent for LRSs, which would appear to confirm that method performance is the primary factor resulting in low recoveries for these compounds in BSSs. Six other parent compounds with low recoveries for BSSs have recoveries greater than 85 percent for LRSs (chlorimuron-ethyl, diflufenzopyr, disulfoton, fenbutatin oxide, nicosulfuron, and orthosulfamuron), indicating the possibility that these compounds are degrading between spiking and sample analysis or that the different (typically lower and less constant) concentrations at which compounds are spiked in BSSs compared with LRSs contribute to the low median recoveries.

As mentioned in the “Data for Laboratory Reagent Spikes, Field Matrix Spikes, and Third-Party Blind Spikes” section of this report, the number of BSSs spiked with relatively high concentrations generally is quite small and does not allow for a robust analysis of any relations between recovery and concentration; however, graphs that show recovery plotted against spiked concentration are available on the QSB Organic Blind Sample Project website (<https://qsb.usgs.gov/OBSP/index.html>). These graphs show recoveries at higher concentrations falling within the range of recoveries at lower concentrations for nearly all pesticides and falling on either side of the line indicating median recovery for most pesticides, which would imply that recoveries do not commonly vary substantially with concentration.

For 23 of the 184 pesticides with sufficient BSS data, the F-pseudosigma values exceed 30 percent (table 7), implying relatively high variability in recovery and, therefore, potentially in reported concentrations. Only naled has an F-pseudosigma value exceeding 30 percent for both LRSs and BSSs, which implies that a factor other than random measurement error, such as degradation, might contribute to high variability for many compounds in BSSs (17 of the compounds with high variability are degradates). However, as with recovery bias, differences in the magnitude and variability of concentrations at which compounds are spiked in BSSs compared with LRSs also could be an important factor for recovery variability. Using RSD as an alternative measure of variability, 76 pesticides would fall outside acceptable limits (table 7). However, most of these compounds have at least one very high recovery result of greater than 250 percent and often greater than 500 percent, which could have a substantial influence on the RSD. It is not known whether these very high recoveries are the result of method performance or reporting issues at the laboratory or if they could reflect some type of error in preparation of the spike mixture or spiked sample.

The graphs in figure 4, as well as the graphs for additional pesticides in Bexfield and others (2020, PlotGroup3), illustrate changes in the typical BSS recovery for individual pesticides with sufficient data for 2013–18.

Similar to results for LRSs, for some pesticides, including imazamox and aldicarb sulfoxide (fig. 4A and B), time periods with a relatively high frequency of particularly low or high BSS recoveries are evident. Again, similar patterns in method performance might be reflected in results reported for groundwater samples and could be an important consideration for certain uses of the groundwater data.

False-negative results for the 184 pesticides that were spiked into at least 10 individual BSSs submitted for analysis during 2013–18 indicate that, when considering only samples that were spiked at a concentration above the maximum RL, false-negative occurrence exceeds 10 percent for 8 compounds (table 5). Of these eight compounds (disulfoton oxon, naled, orthosulfamuron, oxamyl oxime, phorate, phthalazinone, prosulfuron, and terbufos oxon), five also have median recoveries below 70 percent for BSSs, which is consistent with method performance issues and (or) degradation of the compound after spiking that could contribute to the apparent low bias indicated by false-negative results. A total of 104 pesticides have false-negative occurrence exceeding 1 percent when considering only samples that were spiked at a concentration above the maximum RL; this result indicates that these compounds will be detected somewhat less than 99 percent of the time when they are present and their concentration is near the RL (false-negative occurrence would be expected to decrease at higher concentration). It is not known whether this common exceedance of the target of 1 percent false-negative occurrence for the method (Sandstrom and others, 2015) is more reflective of method performance or degradation of compounds between spiking (typically at concentrations only slightly above the RL) and sample analysis. If method performance is the primary factor in relatively high false-negative occurrence for a compound, this generally implies that the RL is too low. For some pesticides, an increase in the RL starting on January 1, 2016, appears to be accompanied by a reduction in false-negative occurrence (for example, see the false-negative results for acetochlor sulfonic acid in fig. 4C and the associated RL increase in table 1).

Field Matrix Spikes

Recovery results for FMSs indicate that most of the 223 pesticides included in the spike solutions used for S2437 meet data quality objectives for both bias and variability even in a groundwater matrix and after days have elapsed between spiking and sample analysis. After data preparation, the number of FMS results available for an individual pesticide ranges from 19 to 81 (table 8, available for download at <https://doi.org/10.3133/sir20205072>). The median recovery for individual pesticides ranges from 0.0 to 167.0 percent (table 8), with 202 pesticides having median recovery between 70 and 130 percent. The median recovery exceeds 130 percent for didealkyltriazine (149.8 percent) and 2-hydroxy-6-ethylamino-4-amino-s-triazine (167.0 percent), potentially as a result of matrix interference and (or) of degradation of

the parent compound in the sample. The median recovery is lower than 70 percent for 19 pesticides. Six of these pesticides (bifenthrin, indoxacarb, lactofen, naled and cis- and trans-permethrin) have median recoveries that are also among the lowest observed for LRSs (below 80 percent), which implies that method performance is an important factor resulting in low recoveries for these pesticides in FMSs. Whereas naled and bifenthrin have relatively low median recoveries for LRSs (71.0 and 74.7 percent, respectively), they have especially low recoveries (below 38 percent) for FMSs (see fig. 5A for side-by-side box plots of recoveries for naled in LRSs and FMSs, which use the same spike solutions). These two pesticides also have especially low recoveries for BSSs; although BSSs use different spike solutions and concentrations, this pattern among spike types could indicate that both method performance and degradation of compounds prior to analysis could contribute to low bias for these compounds (sorption to container walls could also be a factor, particularly for bifenthrin). Considering only pesticides (not degradates), orthosulfamuron and fenbutatin oxide have low recoveries for FMSs and BSSs but a median recovery of at least 85 percent for LRSs, indicating that the low bias shown for FMSs (illustrated relative to LRSs in fig. 5B for orthosulfamuron) and BSSs might be due primarily to degradation. Considering all compounds, 1H-1,2,4-triazole, asulam, and novaluron have substantially lower recoveries for FMSs than for either BSSs or LRSs, indicating the possible importance of matrix effects. Side-by-side box plots of recoveries in FMSs and LRSs are available for all pesticides in Bexfield and others (2020, PlotGroup4).

Of the 118 pesticides with HHBs and results for FMSs, 70 have a median recovery between 90 and 110 percent, implying that reported concentrations for these pesticides typically would have very little bias and could accurately represent any exceedances of HHBs that may occur for groundwater (although spike concentrations typically are well below HHBs). The high median recovery for didealkyltriazine (149.8 percent) indicates a high bias that could cause some reported concentrations to be categorized as being above the HHB when the actual concentration in groundwater is lower than the HHB. Forty-one pesticides with HHBs have a median recovery below 90 percent but above 50 percent, indicating a small to moderate low bias that could cause some reported concentrations to be categorized as being below the HHB when the actual concentration in groundwater is slightly greater than the HHB. Six pesticides with HHBs (1H-1,2,4-triazole, asulam, bifenthrin, cis-permethrin, fenbutatin oxide, and naled) have a median recovery lower than 50 percent and as low as zero percent, indicating that method performance, degradation, and (or) matrix effects could cause substantial underrepresentation of concentrations of these pesticides in groundwater and their relations to HHBs. Most pesticides have at least one recovery value (minimum or maximum) outside the 70–130 percent range; it is not known whether these extreme values are representative of method performance, compound degradation, matrix effects, or a sample error of some type.

FMSs are not ideally suited for evaluation of variability (that is, solely random measurement error) because spatial and (or) temporal differences in environmental sample matrices can substantially affect laboratory performance for some compounds and because some compounds can undergo degradation between spiking and analysis. Nevertheless, examination of F-pseudosigma values for FMSs can provide useful information about which pesticide compounds are more likely than others to be affected by a combination of these factors in addition to random measurement error. For 8 of the 223 pesticides with results for FMSs, F-pseudosigma values exceed 30 percent (table 8), implying relatively high variability in recovery and, therefore, potentially in reported concentrations. Of these eight pesticides, seven have median recoveries outside the acceptable range of 70–130 percent, confirming that the observed variability for these and possibly other compounds likely is affected by degradation and matrix effects and does not reflect method performance alone. Twenty-eight pesticides have RSDs exceeding 30 percent, seven of which also have F-pseudosigma values exceeding 30 percent (table 8).

The graphs in figure 4, as well as the graphs of additional pesticides in Bexfield and others (2020, PlotGroup3), illustrate changes in FMS recovery for individual pesticides with spike results for 2013–18 (the absence of data points for much of 2013 is the consequence of having to drop results for FMSs that used spike lot 91219, as described in the “Data Preparation and Evaluation” section of this report). Temporal patterns in FMS recovery for an individual pesticide, including timing of the occasional occurrence of extreme recovery values, commonly appear to reflect similar patterns in LRS results.

Variability Indicated by Results for Field Replicates

The few pesticides on S2437 with sufficient FR results for evaluation indicate generally low variability in analyte detection and concentration as defined in this section. After data preparation, the number of replicate pairs with results available for an individual pesticide ranges from 136 to 169; however, analysis of variability in analyte detection was performed only for the 13 pesticides that had 10 or more replicate pairs with a detection in at least 1 sample of the pair (table 9). Analysis of variability in analyte concentrations was performed only for the 12 pesticides that had 10 or more replicate pairs with detections in both samples of the pair (table 10).

Among the 13 pesticides for which variability in detection was estimated, 12–51 replicate pairs had a detection in at least 1 sample of the pair (table 9). The mean detection rate for these pesticides ranges from 78.9 to 96.9 percent. The percentage of pairs with inconsistent detections ranges from 6.3 to 42.1 percent when only replicate pairs with at least one detection are considered and ranges from 0.6 to 4.8 percent

when all replicate pairs are considered. To indicate high variability of detection, Martin (2002) used a mean detection rate of 75 percent or less or a percentage of inconsistent replicate sets of 50 percent or more when only replicate pairs with at least one detection were considered. Using those same criteria for the FR data available for the current study, none of the pesticides analyzed by S2437 would be considered to have high variability of detection. Martin (2002) also used a mean detection rate of 90 percent or more or a percentage of inconsistent replicate sets of 25 percent or less when only replicate pairs with at least one detection were considered to indicate low variability of detection. Using these criteria, 2-hydroxy-4-isopropylamino-6-amino-s-triazine, hexazinone, and metolachlor do not have low variability of detection but rather would be considered to have moderate variability. Of these three pesticides, metolachlor is the only one detected in at least 1 percent of FBs and (or) LBs, implying the potential for false-positive results to contribute to a higher variability of detection. Although false-negative rates based on BSSs are acceptable for all three pesticides, matrix interferences or other loss processes could contribute to false-negative results and associated variability in detection in FRs.

If objectives of a groundwater study using S2437 data required results for detection frequency to be highly reproducible, more detailed analysis of variability of detection, such as through calculation of an upper confidence bound on percentages of inconsistent replicate sets by compound, might be needed (Mueller and others, 2015). Also, because variability of detection tends to be substantially higher at low concentrations (less than the RL) and to decrease with increasing concentrations (Martin, 2002), analysis of variability in detection by concentration range for individual compounds might be appropriate. The current dataset does not include enough replicate pairs with at least one detection to perform a robust analysis by concentration range for any individual pesticide. A broad analysis across the 81 pesticides with at least 1 replicate pair having a detection indicates that 49.6 percent of pairs with the minimum reported concentration being lower than the maximum RL have inconsistent detections, compared with 3.3 percent of pairs with the minimum reported concentration being greater than the maximum RL; however, method performance with respect to variability in detection is likely to vary substantially by pesticide (table 9).

The 12 pesticides evaluated for variability in concentration had 11–46 replicate pairs with detections in both samples of the pair. The mean SD and (or) RSD determined for specified concentration ranges for these pesticides by use of the two-range model are presented in table 10. Graphs supporting selection of an appropriate boundary between the concentration ranges listed in table 10 are available in PlotGroup5 of Bexfield and others (2020). Eleven pesticides had sufficient data to calculate a mean SD for lower concentrations; the mean SD for each of these pesticides was quite low at less than 5 ng/L, with the exceptions of deisopropylatrazine (mean SD of 7.6 ng/L) and metolachlor sulfonic acid (mean SD of 17.6 ng/L). Only three pesticides (didealkylatrazine, metolachlor oxanilic acid, and metolachlor

Table 9. Estimated variability in detection of schedule 2437 pesticide compounds based on field replicates, May 2013 through September 2018.

[Values are based on data from U.S. Geological Survey (2019) and published in Bexfield and others (2020). Compounds are sorted by U.S. Geological Survey (USGS) parameter code. Only pesticide compounds having 10 or more replicate pairs with a detection in at least one sample of the pair are included in this table]

Pesticide compound	USGS parameter code	Total number of replicate pairs	Number of replicate pairs with a detection in at least one sample	Number of replicate pairs with inconsistent detections	Mean detection rate (percent)	Of all replicate pairs, percent with inconsistent detections	Of replicate pairs with at least one detection, percent with inconsistent detections
Atrazine	65065	164	45	6	93.3	3.7	13.3
Hexazinone	65085	165	18	7	80.6	4.2	38.9
Metolachlor	65090	165	19	8	78.9	4.8	42.1
Simazine	65105	165	23	4	91.3	2.4	17.4
Prometon	67702	165	19	2	94.7	1.2	10.5
Didealkylatrazine	68547	165	48	3	96.9	1.8	6.3
Deisopropylatrazine	68550	165	28	3	94.6	1.8	10.7
Deethylatrazine	68552	165	49	8	91.8	4.8	16.3
Dechlorometolachlor	68562	165	12	3	87.5	1.8	25.0
Metolachlor oxanilic acid	68650	169	13	1	96.2	0.6	7.7
Metolachlor sulfonic acid	68651	169	51	5	95.1	3.0	9.8
2-Hydroxy-4-isopropylamino-6-amino-s-triazine	68659	165	19	6	84.2	3.6	31.6
2-Hydroxyatrazine	68660	165	14	2	92.9	1.2	14.3

Table 10. Estimated variability in concentration of schedule 2437 pesticide compounds based on field replicates, May 2013 through September 2018.

[Values are based on data from U.S. Geological Survey (2019) and published in Bexfield and others (2020). Compounds are sorted by U.S. Geological Survey (USGS) parameter code. Only pesticide compounds having 10 or more replicate pairs with a detection in both samples of the pair are included in this table. ng/L, nanogram per liter; SD, standard deviation; RSD, relative standard deviation; <, less than; >, greater than]

Pesticide compound	USGS parameter code	Number of replicate pairs with detections in both samples	Concentration range (ng/L)	Number of replicate pairs in concentration range	Statistic	Value	Units
Atrazine	65065	39	<90	35	Mean SD	1.2	ng/L
			>90	4	Mean RSD	Insufficient data	percent
Hexazinone	65085	11	<20	10	Mean SD	0.4	ng/L
			>20	1	Mean RSD	Insufficient data	percent
Metolachlor	65090	11	<30	10	Mean SD	0.9	ng/L
			>30	1	Mean RSD	Insufficient data	percent
Simazine	65105	19	<50	17	Mean SD	2.1	ng/L
			>50	2	Mean RSD	Insufficient data	percent
Prometon	67702	17	<10	10	Mean SD	0.4	ng/L
			>10	7	Mean RSD	Insufficient data	percent
Didealkylatrazine	68547	45	<65	15	Mean SD	4.6	ng/L
			>65	30	Mean RSD	12.5	percent
Deisopropylatrazine	68550	25	<500	24	Mean SD	7.6	ng/L
			>500	1	Mean RSD	Insufficient data	percent
Deethylatrazine	68552	41	<200	36	Mean SD	4.1	ng/L
			>200	5	Mean RSD	Insufficient data	percent
Metolachlor oxanilic acid	68650	12	<100	1	Mean SD	Insufficient data	ng/L
			>100	11	Mean RSD	6.4	percent
Metolachlor sulfonic acid	68651	46	<500	29	Mean SD	17.6	ng/L
			>500	17	Mean RSD	5.2	percent
2-Hydroxy-4-isopropylamino-6-amino-s-triazine	68659	13	<20	12	Mean SD	0.4	ng/L
			>20	1	Mean RSD	Insufficient data	percent
2-Hydroxyatrazine	68660	12	<50	10	Mean SD	0.8	ng/L
			>50	2	Mean RSD	Insufficient data	percent

sulfonic acid) had sufficient data to calculate a mean RSD for higher concentrations. The mean RSD was low at less than 10 percent for metolachlor oxanilic acid and metolachlor sulfonic acid (6.4 percent and 5.2 percent, respectively) and was 12.5 percent for didealkylatrazine. These results for the variability of reported pesticide concentrations could be used to estimate the uncertainty of the concentration measured for a single environmental sample and the associated probability of exceedance of an HHB or to estimate the minimum difference in mean concentrations that can be measured with a certain level of confidence (Mueller and others, 2015).

Overall, the available FR results indicate generally low variability in analyte detection and concentration, meaning that random measurement error has minimal potential to affect the number of groundwater samples that would be classified as having the presence of a pesticide or as having moderate or high concentrations relative to current HHBs. However, these conclusions can be drawn for only a small subset of pesticides that were detected with sufficient frequency in FRs to evaluate variability in detection and concentration. For other pesticides analyzed by S2437, more data would be required to definitively determine if these conclusions would apply, although the consistency demonstrated in reporting of nondetections for replicate pairs for many pesticides provides some evidence of low variability in analyte detection. Additional data on variability in environmental matrices are available from spikes performed during method validation (tables 22 and 23 of Sandstrom and others, 2015).

Implications for Interpretation of Schedule 2437 Pesticide Results for Groundwater

This quality assessment for pesticide results determined using S2437 indicates that measurements produced by the method for most pesticides have bias and variability that can be considered acceptable for many interpretative studies, which could therefore use the results without qualification or censoring. However, the reported data for a subset of pesticides have the potential for unacceptable contamination bias, high or low recovery bias, or high variability as a consequence of method performance and (or) nonlaboratory factors that could preclude their use for certain common objectives or could necessitate adjustment or qualification to meet those objectives; these potential data-quality issues are summarized for all pesticides in table 11 (available for download at <https://doi.org/10.3133/sir20205072>). This section has a basis in the discussion in appendix 3 of Medalie and others (2019) regarding considerations for choosing censoring thresholds to apply to NWQL data for certain study objectives, but this section focuses in particular on quality issues and data interpretation for S2437 results reported for groundwater.

Considerations Regarding Contamination Bias

The evaluation of contamination bias for S2437 pesticides found little evidence of bias based on FBs but indicated the potential for environmentally relevant levels of contamination bias for some pesticides based on LBs and on unspiked BSSs (table 11). Therefore, censoring of some detections in environmental samples might be necessary to avoid an unacceptably high probability of a false-positive result caused by contamination by the processes that affected the LBs and unspiked BSSs. The threshold(s) used for censoring should be carefully considered on the basis of study objectives.

As addressed by Medalie and others (2019), some studies of pesticide occurrence are specifically designed to limit the probability of a false-positive detection to no more than 1 percent, such as the GAMA project described by Fram and Stork (2019). This study objective would minimally require censoring of S2437 results at the DL (that is, classification of reported detections below that concentration as nondetections), regardless of any additional data-quality considerations. Results of the current data-quality assessment indicate that further censoring might be necessary for some pesticides in order not to exceed a 1-percent probability of a false-positive occurrence. Propoxur was the only pesticide with any false-positive results for BSSs analyzed in 2018, but results for unspiked compounds for BSSs analyzed in 2013–18 indicated that 31 pesticides had false-positive occurrence above the maximum DL that exceeded 1 percent (see the “Third-Party Blanks and Spikes” section of this report). Although 16 of these compounds were detected in fewer than 1 percent of LBs, which does not provide clear support for the BSS results and might imply that spiking issues or degradation after spiking could be contributing factors to false-positive occurrence for these compounds, high false-positive occurrence (particularly in combination with common detection in LBs) indicates the need to consider censoring of some detections reported above the DL, at least during certain time periods, to reduce the probability of false-positive results for some compounds.

As discussed in the “Laboratory Blanks” section of this report, more than half of the pesticides on S2437 had a maximum LB detection during 2013–18 that exceeded the maximum DL (table 3); however, the number of exceedances of the maximum DL typically was small for any individual pesticide, and the 90-percent UCL for the 95th percentile of LB concentrations exceeded the maximum DL for any given water year for only 17 of 225 pesticides (table 4). To ensure achieving the desired probability of a false-positive occurrence, examples of alternate censoring thresholds (greater than the DL) that might be considered for pesticides with evidence of unacceptable contamination bias during certain periods of analysis by S2437 include the RL, the UCL for a high percentile of LB concentrations, the maximum LB concentration, or a value that is some multiple of the maximum LB concentration (see app. 3 of Medalie and

others, 2019, for a more detailed discussion of alternate censoring thresholds and references to studies that have used them). As discussed in a subsequent paragraph of this section, censoring thresholds could be applied only during relatively short periods that met a given definition of “episodes” of contamination.

Some studies of pesticide occurrence, generally including NAWQA studies of groundwater or surface water, have an objective of reporting the widest likely occurrence of pesticides in groundwater or surface water, even at trace concentrations. Compared with studies like GAMA, such studies may be more tolerant of a higher probability of false-positive results and less tolerant of a higher probability of false-negative results. As censoring levels increase, the likelihood of failing to identify and report the presence of a compound in the environment (a false-negative result) also increases, which can be an important loss of information for these studies, particularly when comparing concentrations to benchmarks. For the few pesticides that have a water-quality benchmark within one or two orders of magnitude of the DL, an increase in false-negative results may be particularly detrimental to meeting study objectives. Therefore, for studies of this type, it generally is appropriate to use results reported by the laboratory with no additional censoring unless there is evidence of contamination bias that would unacceptably increase the risk of false-positive results. In this case, one or more of the censoring thresholds mentioned in the preceding paragraph could be chosen for application to individual pesticides during certain time periods, with the threshold being selected on the basis of characteristics of the contamination and the desired confidence that contamination of the environmental sample does not exceed that threshold.

As an example of a censoring approach, when quantifiable for an individual compound, the 90-percent UCL for the 95th percentile of LB concentrations during a certain time period could be applied as a censoring level to ensure that, with 90-percent confidence, contamination bias would exceed that concentration in no more than 5 percent of environmental samples. As discussed in the “Laboratory Blanks” section of this report, 62 pesticides on S2437 have a quantifiable 90-percent UCL for the 95th percentile for at least 1 water year, and use of this value as a censoring threshold would not substantially affect the use of reported groundwater results to identify exceedances of the HHB of a pesticide or 10 percent of the HHB of a pesticide. Table 1 of Bexfield and others (2020) includes pesticide results for the groundwater samples collected for NAWQA in 2013–18 as reported by the laboratory and as they would be reported after use of the 90-percent UCL for the 95th percentile of LB concentrations by water year as a censoring threshold (that is, detections at concentrations below this censoring threshold would be considered nondetections at a concentration less than that threshold, and reported as “<” that threshold). To provide a broad illustration of the effects of this censoring approach on that dataset, which has between 857 and 1,934 results per pesticide, the detection frequency for 130 of the 157 pesticides

with at least 1 groundwater detection among all samples in the uncensored dataset did not change as a consequence of censoring. The detection frequency changed by 0.5 percent or more for 3 pesticides: metalaxyl (10 of 31 detections censored, for a decrease in detection frequency of 0.5 percent), metolachlor (19 of 116 detections censored, for a decrease in detection frequency of 1.0 percent), and atrazine (66 of 378 detections censored, for a decrease in detection frequency of 3.5 percent).

For either of the general data objectives discussed above, when detections in LBs are clustered in time, an alternative to using a single censoring level determined from all LBs analyzed during a relatively long period (several months or years) is to define and identify individual contamination episodes. Censoring levels then can be determined for affected compounds and applied to detections that were reported for environmental samples analyzed within these episodes (Medalie and others, 2019). Fram and Belitz (2011), Bexfield and others (2019), Fram and Stork (2019), and Medalie and others (2019) have defined and identified contamination episodes for individual organic compounds by examining the moving average detection frequency in a specified number of LBs (generally 21) and applying censoring when detections (either all detections or detections above a laboratory limit) exceeded a user-defined frequency such as 5 or 10 percent.

For pesticides analyzed by using previous NWQL schedules and investigated by Medalie and others (2019) because of their relatively high detection frequency in LBs, most laboratory contamination was found to be episodic, meaning that it occurred during discrete periods of time, generally between 1 and 8 months in length. For their censoring approach, Fram and Stork (2019) defined contamination episodes during which they applied a censoring threshold at three times the highest concentration detected in an LB during the episode and then calculated a separate high percentile of concentrations for LBs analyzed outside those episodes for consideration as a separate censoring threshold. Laboratory contamination episodes were not defined and identified for the current data-quality assessment, given that different definitions could be appropriate for different study objectives. On the basis of plots of reported LB detections and false-positive results by analysis date, some S2437 pesticides do appear to have discrete episodes of laboratory contamination (although it is possible that some discrete periods of high false-positive occurrence could be associated with unidentified issues with individual BSS spike solutions). However, on the basis of the numbers and temporal patterns of NAWQA groundwater detections, LB detections for most pesticides with substantial numbers of detections in both LBs and NAWQA groundwater samples during similar periods (such as atrazine and metolachlor) generally appear to be spread across several months or across years, rather than occurring in relatively short episodes.

It is useful to note that for S2437 pesticides the observation that FBs show substantially less evidence of contamination bias compared with LBs appears to indicate

that the laboratory practice of set-by-set censoring reduces the risk of false-positive results in field samples (blanks and environmental samples) below the level that LB results alone would imply. Therefore, censoring thresholds applied on the basis of LB results appear likely to be conservative, meaning that they are higher than the concentrations actually necessary to achieve the desired level of confidence regarding the potential contribution of contamination to environmental sample results. However, assuming minimal problems with degradation or with spike solution errors, the relatively high false-positive occurrence rates for some compounds for unspiked BSSs, which should be subject to laboratory censoring, could indicate that set-by-set censoring does not always achieve a low probability of false-positive results.

Considerations Regarding Bias in Recovery

The evaluation of spike samples for S2437 pesticides found that the typical recovery bias for most pesticides in groundwater samples likely is within an acceptable range for most uses of the data. Results for LRSs show that all S2437 pesticides have a median recovery within a range of 70–110 percent in reagent water, with only seven pesticides having a median recovery below 80 percent (table 6). These results indicate that in the absence of degradation or matrix effects little or no adjustment of environmental results would be necessary to provide accurate comparisons of environmental concentrations to benchmarks of interest, such as HHBs. However, as discussed in the “Third-Party Blind Spike Samples” and “Field Matrix Spikes” sections of this report, degradation and (or) matrix effects do appear to adversely affect recovery for several pesticides (table 11), typically resulting in a negative bias; for BSSs, these results might also reflect the generally lower concentrations used relative to other spike types. Relatively high false-negative occurrence rates provide additional confirmation of a negative bias for some pesticides. For pesticides with a particularly strong negative bias indicated for FMSs, actual exceedances of an HHB in groundwater could be substantially underrepresented by the reported concentrations. Therefore, such comparisons should be qualified to express the potential effects of recovery bias on conclusions; the typical recovery could be used to estimate a reasonable upper bound on the actual environmental concentration. Similarly, for the small number of pesticides with the likelihood of a strong positive bias based on recovery (tables 8 and 11), actual exceedances of a benchmark could be overrepresented, and conclusions regarding exceedances should be qualified, possibly with inclusion of an estimate of a reasonable lower bound on the groundwater concentration considering the laboratory measurement.

Graphs of recovery through time for the various types of spikes illustrate that bias can differ depending on the date when a sample was collected and analyzed. These potential differences in bias should be considered when examining changes over time in pesticide results for environmental samples, as concentrations might need to be adjusted to a

uniform recovery value to ensure that any reported changes in concentration reflect environmental conditions rather than laboratory method performance. Bexfield (2008), Martin and others (2009), and Martin and Eberle (2011) discussed the implications of changes in recoveries in LRSs and FMSs for the analysis of temporal trends in NAWQA data for pesticides in groundwater and (or) surface water.

Considerations Regarding Variability

The evaluation of spike samples for S2437 pesticides indicated that the variability in recovery for most pesticides in groundwater samples likely is within an acceptable range for most uses of the data. Results for LRSs indicate that only two pesticides on S2437 have F-pseudosigma values greater than 30 percent for recovery (table 6). These values indicate that random measurement error generally is quite small; therefore, in the absence of degradation or matrix effects, groundwater results are expected to be highly reproducible. However, as discussed in the “Third-Party Blind Spike Samples” section of this report, degradation might adversely affect recovery for some compounds (table 11), increasing variability in results. Differences in the concentrations at which compounds are spiked in BSSs compared with LRSs also could be a factor. For compounds with greater variability, the confidence interval on any individual concentration is larger, which has implications for comparing concentrations to benchmarks and for determining whether two values collected at separate locations or times are truly different with a specified level of confidence.

Detailed analysis of variability using FR data is possible for only a small subset of S2437 pesticides on the basis of the current dataset. However, results indicate generally low variability in analyte detection and concentration for the studied pesticides in groundwater, meaning that random measurement error is unlikely to substantially affect conclusions regarding exceedances of benchmarks or differences in concentrations between samples.

Summary

For studies conducted using results for 225 pesticide compounds (pesticides and pesticide degradates, hereafter referred to as “pesticides”) determined by the U.S. Geological Survey National Water Quality Laboratory schedule 2437 (S2437), including the National Water-Quality Assessment Project groundwater and surface-water studies, it is necessary to assess the ability of reported results to meet data-quality requirements that will allow study objectives to be achieved. This assessment of the quality of S2437 data reported for 2013–18 examined data from field and laboratory quality-control samples, along with third-party performance assessment samples, to estimate and summarize bias and variability and to identify their sources, with an emphasis on implications for the interpretation of pesticide data for groundwater.

Whereas data for field blanks (FBs) show little evidence of substantial contamination bias for S2437 pesticides, data for laboratory blanks (LBs) and for pesticides that were not spiked into blind spike samples (BSSs) indicate that groundwater results for some pesticides have the potential to be substantially affected by contamination from laboratory processes. The observation of less evidence of contamination bias in FBs relative to LBs could largely be the consequence of laboratory data-reporting practices, which utilize detections in LBs to censor results in associated field samples (including blanks and environmental samples) when relative concentrations indicate that a result could have a substantial contribution from laboratory contamination. However, relatively high false-positive occurrences for some unspiked compounds for BSSs indicate that laboratory censoring might not always achieve a low probability of false-positive results.

Censoring of some detections in environmental samples based on LB and unspiked BSS results might be necessary or desirable to avoid an unacceptably high likelihood of a false-positive result caused by contamination by the processes that affected these samples. Plots of LB and BSS data by analysis date indicate that detections and (or) concentrations can vary substantially with time; therefore, the 90-percent upper confidence limits (UCLs) on the 90th and 95th percentiles of LB concentration for each pesticide were calculated by water year. The 90-percent UCL for the more conservative 95th percentile equals or exceeds the minimum reported groundwater concentration in at least 1 water year for 28 pesticides, indicating that at least some concentrations reported for these pesticides could result from laboratory contamination bias rather than being representative of groundwater conditions. This UCL exceeds the maximum laboratory detection limit during at least 1 water year for 17 pesticides: 2,4-D, 2-hydroxy-4-isopropyl-6-ethylamino-s-triazine, ametryn, atrazine, cis-permethrin, diazinon, diuron, fenbutatin oxide, fipronil sulfide, imazaquin, metalaxyl, metconazole, pendimethalin, propiconazole, propoxur, pyriproxyfen, and trans-permethrin. In addition, this UCL exceeds the maximum laboratory reporting limit during 1 water year each for ametryn, atrazine, and diazinon. Laboratory contamination bias in 95 percent of groundwater samples, with 90-percent confidence, should not substantially affect the suitability of reported environmental concentrations for any pesticide for comparison with its human-health benchmark (HHB) or 10 percent of its HHB.

Whereas data available for blind blank samples analyzed in 2018 indicate that only propoxur had any false-positive results, data for unspiked compounds for BSSs analyzed in 2013–18 indicate that the false-positive rates for 31 pesticides at concentrations greater than the maximum detection limit exceeded 1 percent. Although about half of these compounds lack substantial supporting evidence of contamination bias based on LB or FB detections, indicating that spiking issues or degradation of parent compounds within the BSSs might be a contributing factor to some false-positive results, these data indicate that contamination bias might

contribute to detections reported for some pesticides in environmental samples analyzed during a similar period. Data for BSSs that were spiked at concentrations above the maximum reporting limit indicate that false-negative rates for eight pesticides (disulfoton oxon, naled, orthosulfamuron, oxamyl oxime, phorate, phthalazinone, prosulfuron, and terbufos oxon) exceed 10 percent; therefore, low bias could affect results reported for these pesticides in environmental samples analyzed during a similar period.

Data for laboratory reagent spikes show little evidence for unacceptable recovery bias for S2437 pesticides, but field matrix spikes (with supporting evidence from BSSs) indicate that degradation and (or) matrix effects could result in substantial low bias for groundwater results for some pesticides. Two compounds (didealkyltriazine and 2-hydroxy-6-ethylamino-4-amino-s-triazine) have median recoveries in field matrix spikes near or greater than 150 percent, indicating a high bias that could cause some reported concentrations to be categorized as being above the HHB when the actual concentration in groundwater is lower than the HHB. Forty-one pesticides with HHBs have a median recovery between 50 and 90 percent, indicating a small to moderate low bias that could affect comparison of reported concentrations with HHBs. Six pesticides with HHBs (1H-1,2,4-triazole, asulam, bifenthrin, cis-permethrin, fenbutatin oxide, and naled) have a median recovery between zero and 50 percent, indicating that method performance, degradation, and (or) matrix effects could cause substantial underrepresentation of concentrations of these pesticides in groundwater and their relations to HHBs. Plots of data for all spike types show clear changes in the typical recovery with time for some pesticides, which would require further examination for their potential effects on evaluation of temporal trends in environmental concentrations.

Data for laboratory reagent spikes indicate that nearly all S2437 pesticides have acceptable variability resulting from random measurement error. Only two pesticides (fenbutatin oxide and naled) have F-pseudosigma values greater than 30 percent for recovery, which implies the potential for relatively high variability in reported concentrations. These 2 pesticides are among the 19 pesticides that have routinely been reported for S2437 since method implementation with a value qualifier code of “m,” meaning that results determined for the compound are highly variable. Data for BSSs show relatively high variability for a greater number of pesticides; these results could reflect the influence of degradation in addition to laboratory method performance, although differences in the concentrations at which compounds are spiked in BSSs compared with laboratory reagent spikes also could be a factor.

Detailed analysis of variability using field replicate data is possible for only a small subset of S2437 pesticides on the basis of the current dataset. Results indicate generally low variability in analyte detection and concentration for these pesticides in groundwater, meaning that random measurement error is unlikely to substantially affect conclusions regarding exceedances of benchmarks or differences in concentrations between samples.

Overall, results of this data-quality assessment indicate that the S2437 pesticide method provides measurements with bias and variability that would be considered acceptable for many interpretative studies, which could therefore use the results without qualification or censoring. However, the reported data for a subset of pesticides have the potential for unacceptable contamination bias, high or low recovery bias, or high variability as a consequence of method performance and (or) nonlaboratory factors that could preclude their use for certain common objectives or could necessitate adjustment or qualification to meet those objectives. In particular, as indicated by evaluation of results for LBs over days to years, censoring of some environmental detections reported by the laboratory might be necessary to avoid an unacceptably high probability of a false-positive result caused by contamination from laboratory processes and (or) carryover. However, it should be noted that laboratory censoring practices appear likely to reduce the risk of false-positive results in environmental samples below the level that LB results alone would imply. Also, as censoring levels increase, the likelihood of failing to identify and report the presence of a compound in the environment (a false-negative result) also increases, which can be an important loss of information for some studies, particularly when comparing concentrations to benchmarks. Relatively strong positive or negative recovery bias for some pesticides because of method performance, compound degradation, and (or) matrix effects could necessitate qualifications of conclusions regarding benchmark exceedances. In addition, relatively high variability for some pesticides as a consequence of these factors could have implications for comparing concentrations to benchmarks and for determining whether two values collected at separate locations or times are truly different with a specified level of confidence.

Acknowledgments

This effort was made possible by the hard work of numerous U.S. Geological Survey (USGS) field crews and by USGS National Water Quality Laboratory personnel who analyzed samples and assisted in quality-control evaluations.

References Cited

- Altman, N.S., 1992, An introduction to kernel and nearest-neighbor nonparametric regression: *The American Statistician*, v. 46, no. 3, p. 175–185.
- Arnold, T.L., Bexfield, L.M., Musgrove, M., Lindsey, B.D., Stackelberg, P.E., Barlow, J.R., DeSimone, L.A., Kulongoski, J.T., Kingsbury, J.A., Ayotte, J.D., Fleming, B.J., and Belitz, K., 2017, Datasets from groundwater-quality data from the National Water-Quality Assessment Project, January through December 2014 and select quality-control data from May 2012 through December 2014: U.S. Geological Survey data release, <https://doi.org/10.5066/F7W0942N>.
- Arnold, T.L., Bexfield, L.M., Musgrove, M., Stackelberg, P.E., Lindsey, B.D., Kingsbury, J.A., Kulongoski, J.T., Belitz, K., and Sharpe, J.B., 2018, Datasets from groundwater-quality and select quality-control data from the National Water-Quality Assessment Project, January through December 2015 and previously unpublished data from 2013–2014: U.S. Geological Survey data release, <https://doi.org/10.5066/F7XK8DHK>.
- Arnold, T.L., Sharpe, J.B., Bexfield, L.M., Musgrove, M., Erickson, M.L., Kingsbury, J.A., Degnan, J.R., Tesoriero, A.J., Kulongoski, J.T., and Belitz, K., 2020, Datasets from groundwater-quality and select quality-control data from the National Water-Quality Assessment Project, January through December 2016, and previously unpublished data from 2013 to 2015: U.S. Geological Survey data release, <https://doi.org/10.5066/P9W4RR74>.
- Arnold, T.L., DeSimone, L.A., Bexfield, L.M., Lindsey, B.D., Barlow, J.R., Kulongoski, J.T., Musgrove, M., Kingsbury, J.A., and Belitz, K., 2016, Groundwater quality data from the National Water-Quality Assessment Project, May 2012 through December 2013: U.S. Geological Survey data release, <http://dx.doi.org/10.5066/F7HQ3X18>.
- ASTM International, 2007, Standard practice for 99 %/95 % interlaboratory detection estimate (IDE) for analytical methods with negligible calibration error: West Conshohocken, Pa., ASTM International, ASTM D6091–07, 13 p., accessed August 12, 2019, at <http://www.astm.org/Standards/D6091.htm>.
- ASTM International, 2010, Standard practice for performing detection and quantitation estimation and data assessment utilizing DQCALC software, based on ASTM practices D6091 and D6512 of Committee D19 on water: West Conshohocken, Pa., ASTM International, ASTM D7510–10, 2 p., accessed August 12, 2019, at <http://www.astm.org/Standards/D7510.htm>.
- Belitz, K., Jurgens, B., Landon, M.K., Fram, M.S., and Johnson, T., 2010, Estimation of aquifer scale proportion using equal area grids—Assessment of regional scale groundwater quality: *Water Resources Research*, v. 46, accessed June 4, 2020, at <https://doi.org/10.1029/2010WR009321>.
- Bender, D.A., Zogorski, J.S., Mueller, D.K., Rose, D.L., Martin, J.D., and Brenner, C.K., 2011, Quality of volatile organic compound data from groundwater and surface water for the National Water-Quality Assessment Program, October 1996–December 2008: U.S. Geological Survey Scientific Investigations Report 2011–5204, 128 p., accessed August 19, 2013, at <https://doi.org/10.3133/sir20115204>.
- Bexfield, L.M., 2008, Decadal-scale changes of pesticides in ground water of the United States, 1993–2003: *Journal of Environmental Quality*, v. 37, no. S5, p. S-226–S-239.

- Bexfield, L.M., Sandstrom, M.W. and Beaty, Delicia, 2020, Field, laboratory, and third-party data for assessment of the quality of pesticide results reported by the National Water Quality Laboratory for groundwater samples collected by the National Water-Quality Assessment Project, 2013–18: U.S. Geological Survey data release, <https://doi.org/10.5066/P90BFKA4>.
- Bexfield, L.M., Toccalino, P.L., Belitz, K., Foreman, W.T., and Furlong, E.T., 2019, Hormones and pharmaceuticals in groundwater used as a source of drinking water across the United States: *Environmental Science & Technology*, v. 53, p. 2950–2960, accessed March 19, 2019, at <https://doi.org/10.1021/acs.est.8b05592>.
- Brown, L.D., Cai, T.T., and DasGupta, A., 2001, Interval estimation for a binomial proportion: *Statistical Science*, v. 16, no. 2, p. 101–133.
- Fram, M.S., and Belitz, K., 2011, Occurrence and concentrations of pharmaceutical compounds in groundwater used for public drinking-water supply in California: *Science of the Total Environment*, v. 409, no. 18, p. 3409–3417, accessed February 10, 2014, at <https://doi.org/10.1016/j.scitotenv.2011.05.053>.
- Fram, M.S., Olsen, L.D., and Belitz, K., 2012, Evaluation of volatile organic compound (VOC) blank data and application of study reporting levels to groundwater data collected for the California GAMA Priority Basin Project, May 2004 through September 2010: U.S. Geological Survey Scientific Investigations Report 2012–5139, 94 p., accessed August 19, 2013, at <https://doi.org/10.3133/sir20125139>.
- Fram, M.S., and Stork, S.V., 2019, Determination of study reporting limits for pesticide constituent data for the California Groundwater Ambient Monitoring and Assessment Program Priority Basin Project, 2004–18, Part 1—National Water Quality Laboratory schedules 2003, 2032, or 2033, and 2060: U.S. Geological Survey Scientific Investigations Report 2019–5107, 129 p., accessed January 8, 2020, at <https://doi.org/10.3133/sir20195107>.
- Helsel, D.R., 2005, *Nondetects and data analysis*: Hoboken, N.J., John Wiley & Sons, Inc., 250 p.
- Herrmann, E., 2016, *lokern*—Kernel regression smoothing with local or global plug-in bandwidth: R package version 1.1–8, accessed January 4, 2020, at <https://cran.r-project.org/web/packages/lokern/index.html>.
- Hoaglin, D.C., Mosteller, F., and Tukey, J.W., eds., 1983, *Understanding robust and exploratory data analysis*: New York, John Wiley, 447 p.
- Martin, J.D., 2002, Variability of pesticide detections and concentrations in field replicate water samples collected for the National Water-Quality Assessment Program, 1992–97: U.S. Geological Survey Water-Resources Investigations Report 2001–4178, 84 p., accessed October 2, 2015, at <https://doi.org/10.3133/wri20014178>.
- Martin, J.D., and Eberle, M., 2011, Adjustment of pesticide concentrations for temporal changes in analytical recovery, 1992–2010: U.S. Geological Survey Data Series 630, 11 p., 5 apps., accessed December 22, 2016, at <https://doi.org/10.3133/ds630>.
- Martin, J.D., Norman, J.E., Sandstrom, M.W., and Rose, C.E., 2017, A field study of selected U.S. Geological Survey analytical methods for measuring pesticides in filtered stream water, June–September 2012: U.S. Geological Survey Scientific Investigations Report 2017–5049, 106 p., accessed October 24, 2017, at <https://doi.org/10.3133/sir20175049>.
- Martin, J.D., Stone, W.W., Wydoski, D.S., and Sandstrom, M.W., 2009, Adjustment of pesticide concentrations for temporal changes in analytical recovery, 1992–2006: U.S. Geological Survey Scientific Investigations Report 2009–5189, 23 p., 9 apps., accessed December 22, 2016, at <https://doi.org/10.3133/sir20095189>.
- Medalie, L., Sandstrom, M.W., Toccalino, P.L., Foreman, W.T., ReVello, R.C., Bexfield, L.M., and Riskin, M.L., 2019, Use of set blanks in reporting pesticide results at the U.S. Geological Survey National Water Quality Laboratory, 2001–15: U.S. Geological Survey Scientific Investigations Report 2019–5055, 147 p., accessed August 9, 2019, at <https://doi.org/10.3133/sir20195055>.
- Mueller, D.K., Schertz, T.L., Martin, J.D., and Sandstrom, M.W., 2015, Design, analysis, and interpretation of field quality-control data for water-sampling projects: U.S. Geological Survey Techniques and Methods, book 4, chap. C4, 54 p., accessed July 31, 2015, at <https://doi.org/10.3133/tm4C4>.
- Mueller, D.K., and Titus, C.J., 2005, Quality of nutrient data from streams and ground water sampled during water years 1992–2001: U.S. Geological Survey Scientific Investigations Report 2005–5106, 27 p., accessed April 9, 2019, at <https://doi.org/10.3133/sir20055106>.
- Norman, J.E., Toccalino, P.L., and Morman, S.A., 2018, *Health-Based Screening Levels for evaluating water-quality data* (2d ed.): U.S. Geological Survey web page, accessed April 6, 2020, at <https://doi.org/10.5066/F71C1TWP>.
- R Core Team, 2019, *R—A language and environment for statistical computing*: Vienna, Austria, R Foundation for Statistical Computing, accessed April 1, 2015, at <https://www.r-project.org/>.
- Sandstrom, M.W., Kanagy, L.K., Anderson, C.A., and Kanagy, C.J., 2015, Determination of pesticides and pesticide degradates in filtered water by direct aqueous-injection liquid chromatography-tandem mass spectrometry: U.S. Geological Survey Techniques and Methods, book 5, chap. B11, 54 p., accessed June 8, 2016, at <https://doi.org/10.3133/tm5B11>.

- Shoda, M.E., Nowell, L.H., and Bexfield, L.M., 2017a, National Water-Quality Assessment Project replicate surface water and groundwater pesticide data analyzed by the USGS National Water Quality Laboratory schedule 2437, water years 2013–15: U.S. Geological Survey data release, accessed August 8, 2019, at <https://doi.org/10.5066/F75H7DS8>.
- Shoda, M.E., Nowell, L.H., Bexfield, L.M., Sandstrom, M.W., and Stone, W.W., 2017b, Recovery data for surface water, groundwater and lab reagent samples analyzed by the USGS National Water Quality Laboratory schedule 2437, water years 2013–15: U.S. Geological Survey data release, accessed August 8, 2019, at <https://doi.org/10.5066/F7QZ28G4>.
- Shoda, M.E., Nowell, L.H., Stone, W.W., Sandstrom, M.W., and Bexfield, L.M., 2018, Data analysis considerations for pesticides determined by National Water Quality Laboratory schedule 2437: U.S. Geological Survey Scientific Investigations Report 2018–5007, 458 p., accessed April 3, 2018, at <https://doi.org/10.3133/sir20185007>.
- U.S. Environmental Protection Agency [EPA], 2017, Human health benchmarks for pesticides: U.S. Environmental Protection Agency web page, accessed August 15, 2019, at <https://iaspub.epa.gov/apex/pesticides/f?p=HHBP:home>.
- U.S. Environmental Protection Agency [EPA], 2018, National Primary Drinking Water Regulations: U.S. Environmental Protection Agency web page, accessed August 15, 2019, at <https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations>.
- U.S. Geological Survey [USGS], 2003, Aquifers—Map of the principal aquifers of the United States: U.S. Geological Survey Groundwater Information web page, accessed August 8, 2019, at <https://water.usgs.gov/ogw/aquifer/map.html>.
- U.S. Geological Survey [USGS], 2005, Changes in reporting levels and data qualifiers for selected pesticides and degradation products in schedule 2060: U.S. Geological Survey, National Water Quality Laboratory Technical Memorandum 05.03, 8 p., accessed August 22, 2019, at http://wwwnwql.cr.usgs.gov/tech_memos/nwql.2005-03.pdf.
- U.S. Geological Survey [USGS], 2007, Removal of selected pesticides and degradation products from schedule 2060: U.S. Geological Survey, National Water Quality Laboratory Technical Memorandum 2007.02, 8 p., accessed August 22, 2019, at http://wwwnwql.cr.usgs.gov/tech_memos/nwql.2007-02.pdf.
- U.S. Geological Survey [USGS], 2011, Application of the result-level ‘v’ value qualifier code and ‘E’ remark code to selected organic results reported by the National Water Quality Laboratory (NWQL): U.S. Geological Survey Office of Water Quality Technical Memorandum 2012.01, 4 p., accessed August 12, 2019, at <https://water.usgs.gov/admin/memo/QW/qw12.01.pdf>.
- U.S. Geological Survey [USGS], 2015, Changes to National Water Quality Laboratory (NWQL) procedures used to establish and verify laboratory detection and reporting limits: U.S. Geological Survey, National Water Quality Laboratory Technical Memorandum 15.02, 21 p., accessed August 12, 2019, at https://nwql.usgs.gov/Public/tech_memos/nwql.2015-02.pdf.
- U.S. Geological Survey [USGS], 2019, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed May 7, 2019, at <https://doi.org/10.5066/F7P55KJN>.
- U.S. Geological Survey [USGS], variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A10, accessed March 4, 2019, at <http://pubs.water.usgs.gov/twri9A>.
- Wieben, C.M., 2019, Pesticide recovery data for surface-water and lab reagent samples analyzed by the USGS National Water Quality Laboratory schedule 2437, water years 2016–17: U.S. Geological Survey data release, accessed September 16, 2019, at <https://doi.org/10.5066/P93MWMVF>.

For more information about this publication, contact

Director, [New Mexico Water Science Center](#)
U.S. Geological Survey
6700 Edith Blvd. NE
Albuquerque, NM 87113

For additional information visit
<https://www.usgs.gov/centers/nm-water>

Publishing support provided by
Lafayette Publishing Service Center

