# A meta-analysis of the impacts of best management practices on nonpoint source pollutants.

**Michael Schramm [1*], Duncan Kikoyo [1], Janelle Wright [1], Shubham Jain [1]**

[1] *Texas Water Resources Institute, Texas A&M AgriLife Research, College Station, Texas, USA*

Correspondence*:
Michael Schramm
michael.schramm@ag.tamu.edu

## ABSTRACT

Abstract length and content varies depending on article type. Refer to http://www.frontiersin.org/about/AuthorGuidelines for abstract requirement and length according to article type.

**Keywords: best management practice, water quality, nonpoint source pollution, fecal indicator bacteria, nutrients**

## 1 INTRODUCTION

In the United States (U.S.), major progress towards water quality goals have been achieved through the Clean Water Act of 1972. This progress has been largely attributed to investments and reductions in point source discharges while reductions in nonpoint source pollutants remain a substantial challenge (Benham et al., 2008; National Research Council, 2001; Schramm et al., 2022). Increased pollutant loads and pollutant concentrations in runoff are a particular challenge when dealing with nonpoint source pollution resulting from land use change. The impacts of land use change on hydrology and water quality are well established (Allan, 2004; Carpenter et al., 1998; Bernhardt et al., 2008; Carey et al., 2013; Freeman et al., 2019). Nonpoint source driven fecal indicator bacteria (FIB), nitrogen, phosphorus, and suspended sediment remain major causes of water quality impairments in U.S. rivers and streams despite decades of work. In 2017, the Environmental Protection Agency (EPA) estimated 41% or more of the nation's rivers and streams rated poorly for biological condition due to excess nitrogen or phosphorus (EPA, 2017). FIB remains the leading cause of water body impairment on the Clean Water Act 303(d) list in the United States (EPA, 2017).

Best management practices (BMPs) have been the primary suite of tools for addressing nonpoint source pollution. BMPs are structural or non-structural controls used to mitigate the effects of increased runoff volume, pollutant loads, or pollutant concentrations emanating from diffuse nonpoint sources. BMPs control the delivery of pollutants through a few possible mechanisms. Structural BMPs reduce and retard total volume of runoff, thus reducing both the volume of water and pollutant load. Structural BMPs may also provide a mechanism for physical, chemical, or biological removal of pollutant constituents suspended or dissolved in runoff. Non-structural BMPs (such as nutrient management or livestock management) are utilized to reduce the generation of pollutant runoff by avoiding pollutant generation during critical periods.

Practitioners rely extensively on mechanistic models to plan and evaluate BMP scenarios and resulting water quality. Lintern et al. (2020) found 43% of reviewed BMP effectiveness studies relied completely on modeled outputs, with modeled outputs almost always predicting water quality improvements following BMP implementation. However, field studies are much more likely to demonstrate mixed results including net releases of pollutants under certain conditions (Lintern et al., 2020; Liu et al., 2017). The disconnect between modeled outcomes and field studies might be attributed to (1) overly simplified or incorrect estimates of parameters that represent management practices (Ullrich and Volk, 2009; Fu et al., 2019; Lintern et al., 2020), (2) failure to incorporate uncertainty in estimates (Tasdighi et al., 2018; Fu et al., 2019; Lintern et al., 2020), and (3) assumption of static performance over time (Meals et al., 2010; Liu et al., 2017; Fu et al., 2019).

A source of uncertainty comes from the substantial variability of performance metrics reported in empirical BMP studies (Lintern et al., 2020). Previous empirical reviews generally describe high variability and uncertainty in nitrogen and phosphorus removal and consistent reduction in total suspended sediment concentrations across BMP types (Lintern et al., 2020; Liu et al., 2017; Koch et al., 2014; Clary et al., 2011; Barrett, 2008; Grudzinski et al., 2020). The review literature on the effects of BMPs on FIBs are sparse but generally find extremely high variance in performance across BMPs (Clary et al., 2011; Grudzinski et al., 2020). Accurately characterizing BMP treated runoff quality is challenging due to the high variability in BMP performance. However, influent water quality, seasonality, and BMP age provide some explanatation to highliy heterogenous BMP efficiency data (Barrett, 2005; Liu et al., 2017).
A recent meta-analysis demonstrated the further linkages between local climatic conditions and BMP performance on nitrogen and phosphorus concentrations (Horvath et al., 2023).

## 2 METHODS

We conducted a systematic review of recent (2000-2022) literature to compile field studies documenting the effectiveness of best management practices on fecal indicator bacteria and nutrients concentrations. The systematic review followed guidance provided in the Collaboration for Environmental Evidence systematic review guidelines (Collaboration for Environmental Evidence, 2018). In order to maximize the number of studies included in the review, we included both peer-reviewed studies and unpublished white papers to reduce potential bias against negative results. The inclusion criteria filtered out (1) non-field studies, (2) modelling results, (3) studies that did not evaluate specific BMPs, (4) studies conducted outside of the U.S. or published in a language other than English. We ran search queries in Texas A&M Library Catalog, Web of Science, and Google Scholar. Although results from Google Scholar are not always replicable, we utilized the service to maximize search results for studies not published in academic journals and presumably increase the chance of identifying studies with negative effects. Fecal indicator bacteria study seraches included the following query: “fecal indicator bacteria” OR “E. coli” OR “Escherichia coli” OR “enterococci” OR “enterococcus” AND “best management practices” OR “BMPs” AND “effectiveness” OR “performance”. Nutrient BMP studies utilized a similar query: “nutrient” OR “nitrogen” OR “phosphorus” OR “sediment” OR “TSS” AND “best management practices” OR “BMPs” AND “effectiveness” OR “performance”.

Results from each database were first filtered to remove duplicates. After removal of duplicates, each member of the research team ($n = 4$) was assigned a subset of studies to evaluate if they should be included. **include supplementary tables with inclusion criteria and data variables** Each study was reviewed by two team members and differences in opinion were collectively discussed and agreed upon before progressing. The remaining studies were split among team members for data extraction, again with at least

**Table 1.** Criteria applied for including or excluding studies within the review database.

| Attribute | Inclusion Criteria | Excluision Criteria |
|---|---|---|
| Study type | Journal articles, book chapters, conference papers, unpublished research reports, thesis and dissertations, organizational and agency white papers. | Synopsis or review studies, peports with reductions based on modeled or other estimated reductions (e.g. TMDLs, watershed plans, or modelling studies. |
| Outcomes | Field studies with measured effects on fecal indicator bacteria or nutrient concentrations. | Studies not explicitly linking reductions to a specific BMP or insufficient information to quantify reductions. |
| Geographical context | Studies conducted within the United States. | Studies outside of the United States. |
| Timeframe | Studies published from 2000 through 2022. | Studies published prior to 2000 or after 2022. |

two team members reviewing each study. If data was provided in figures, the data was extracted with the WebPlotDigitizer tool (Rohatgi, 2022).

## Effect size calculation

Prior BMP reviews routinely report the effect of BMPs using efficiency values (see Agouridis et al., 2005; Kroger et al., 2012; Liu et al., 2015; Simpson and Weammert, 2009). Efficiency values are calculated as:

$$\text{Efficiency} = \frac{x_{control} - x_{experiment}}{x_{control}} \times 100,$$

where $x_{control}$ is the pre-treatment or control pollutant concentration and $x_{experiment}$ is the pollutant concentration measured after the BMP intervention. There are several statistical shortcomings (distributional asymetry, skewness, and non-additive properties) when using efficiency to estimate overall effect sizes (Nuzzo, 2018; Cole and Altman, 2017). In comparision, the log ratio of means (*ROM*) provide preferable statistical properties for regression analysis (Osenberg et al., 1997; Hedges et al., 1999). *ROM* quantifies the difference in means between the control and experimental group (Hedges et al., 1999):

$$ROM_i = ln\left(\frac{x_{i,control}}{x_{i,experiment}}\right) = ln(x_{i,control}) - ln(x_{i,experiment}),$$

where $x_{i,control}$ and $x_{i,experiment}$ are the mean pollutant concentrations for experiment $i$. The statistical properties of *lnRR* (normal distribution around zero and additive properties) are preferable to the more commonly reported efficiency metric (Osenberg et al., 1997; Hedges et al., 1999). Use of *ROM* required the exclusion of studies that only provided measures of BMP efficiency and not the underlying data used to derive the metric.

## 2.1 Statistical models

We used the “rma.mv” function in the *metafor* R package to fit multilevel random effects linear models with *ROM* as the effect variable (Viechtbauer, 2010; R Core Team, 2023). Our models specified a nested random effects term accounting for heterogeneity between effect sizes from the same study and for

**Table 2.** Study and effect variables extracted for review.

| Variable | Description |
|---|---|
| Publication Year | Year the study was published |
| Parameter | The specific pollutant measured. |
| Runoff source | Dominant source of runoff (rop fields, livestock pasture, commerical, residential |
| Source type | Major categorization of runoff source: agricultural or urban |
| BMP | BMP evaluated |
| BMP Classification | BMP description based on NRCS conservation practice standards and EPA BMP fact sheets |
| BMP Category | BMP categorization based on structural or management |
| BMP Subcategory | BMP subcategorization based on pollutant removal processes |
| Study scale | Spatial scale of the study area (”lot/field”, ”community”, ”watershed”) |
| Location | Location name used in the study description |
| State | State where the study was conducted |
| Study area | Drainage area in hectares |
| Longitude | Approximated or reported latitude coordinate |
| Latitude | Approximated or reported longitude coordinate |
| Study years | Year or years when data were collected |
| N control | Number of control measurements |
| N experiment | Number of experiemental measurements |
| X control | Mean concentration for control measurements |
| X experiment | Mean concentration for experiemental measurements |
| SE control | Standard error of control measurements |
| SE experiment | Standard error of experimental measurements |
| Minimum control | Minimum control measurement |
| Minimum experiment | Minimum experiment measurement |
| Maximum control | Maximum control measurement |
| Maximum experiment | Maximum experiment measurement |
| SD control | Standard deviation of control measreuments |
| SD experiment | Standard deviation of experimental measurements |
| Units | Units reported by the study |
| Percent reduction | BMP efficiency for studies that only reported efficiency |

heterogeneity between studies. A key feature of meta-analysis is the weighting of effects using sampling variance of individual effect sizes. 45% of fecal indicator bacteria, 68% of TN, 68% of TP, and 70% of TSS effect sizes were missing standard deviations required to estimate sampling variance. Removal of studies due to missing variance information can introduce substantial bias (Kambach et al., 2020), so we imputed missing standard deviations using random forest based multivariate imputation with chained equations using the *mice* R package (Buuren and Groothuis-Oudshoorn, 2011). Sampling variance was estimated with the variance estimator described in Doncaster et al. to reduce bias in small sample studies:

$$v(ROM) = \frac{\sum_{i=1}^{K} (CV^2_{control,i})/K}{n_{control}} + \frac{\sum_{i=1}^{K} (CV^2_{experiment,i})/K}{n_{experiment}},$$

**Table 3.** Summary table of moderator effects for performance of BMPs on fecal indicator bacteria concentrations.

| Moderator | Estimate ($lnRR$) | 95% CI | SE | $T$-statistic | df | $p$-value |
|---|---|---|---|---|---|---|
| Intercept | 0.21 | [-2.07,2.48] | 1.06 | 0.19 | 14 | 0.85 |
| **BMP Subcategories** | | | | | | |
| Filtration | -0.39 | [-6.42,5.63] | 2.81 | -0.14 | 14 | 0.89 |
| Infiltration | -1.24 | [-3.50,1.01] | 1.13 | -1.10 | 64 | 0.28 |
| Livestock | -1.18 | [-3.25,0.89] | 0.96 | -1.23 | 14 | 0.24 |
| Treatment | -1.56 | [-2.63,-0.49] | 0.53 | -2.92 | 64 | $<$0.01 |
| log(Concentration) | 0.20 | [0.084,0.32] | 0.059 | 3.42 | 65 | $<$0.01 |

$I^2_{\text{total}}$=21.48, $I^2_{\text{study}}$=0, $I^2_{\text{effect}}$=21.48; $R^2_{\text{marginal}}$=0.46

where *v* represents the sampling variance, $CV_{control,i}$ and $CV^2_{experiment,i}$ are the coefficients of variation from the *i*th study for studies 1, 2, . . . , *K*.

Our initial models included log transformed pre-treatment pollutant concentration, study scale (plot/field, community, watershed), study length ($n$ years), BMP subcategory (), and pollutant concentration:BMP subcategory interactions were included as fixed effect terms. We used an information-theoretic approach to select the most parsimonious model from the subset of candidate models based on corrected Akaike information criterion ($AIC_c$) estimated with maximum likelihood (Cinar et al., 2021). The final model was selected from candidate models, which included all combination and subsets of the full model, by eliminating candidate models until the $\Delta AIC_c$ reached $\leq 2$ (Burnham et al., 2011). We estimated regression coefficients of the selected model using restricted maximum likelihood. Relative heterogeneity between and within studies were calculated using the $I^2$ metric described in Nakagawa and Santos (2012). Marginal $R^2$ was used to describe the amount of variance explained by fixed effects (Nakagawa and Schielzeth, 2013).

We tested for evidence of publication bias, in the form of small study effect, by using the extension of Egger's regression applied to the multilevel model framework that included adjusted sampling error as a moderator (Nakagawa et al., 2023). We did not identify evidence of publication bias in the surveyed studies for fecal indicator bacteria ($Z$ = -1.17, $p$ = 0.25), total nitrogen, total phosphorus, or total suspended solids; therefore, adjustments for publication bias were not included in the final models. We conducted a sensitivity analysis of the robustness of overall effect sizes to individual studies using leave-one-out analysis (Nakagawa et al., 2023). Outlier studies identified in the sensitivity analysis are highlighted in Table X and were removed prior to fitting the final models.

# 3 RESULTS

## 3.1 Summary of BMP literature

## 3.2 Regression results

Some evidence of a relationship between input concentrations and fecal indicator bacteria reductions. No evidence strong evidence that the type of BMP changed the overall relationship.

**Table 4.** Summary table of moderator effects for performance of BMPs on nitrogen removal.

| Moderator | Estimate (*lnRR*) | 95% CI | SE | $T$-statistic | df | $p$-value |
|---|---|---|---|---|---|---|
| Intercept | 0.95 | [0.38,1.52] | 0.28 | 3.41 | 28 | <0.01 |
| log(Concentration) | -0.034 | [-0.14,0.071] | 0.053 | -0.64 | 124 | 0.525 |
| Adjusted sampling error | -1.54 | [-3.30,0.22] | 0.89 | -1.74 | 124 | 0.085 |

$I^2_{\text{total}}$=96.57, $I^2_{\text{study}}$=52.45, $I^2_{\text{effect}}$=44.12; $R^2_{\text{marginal}}$=0.05

**Table 5.** Summary table of moderator effects for performance of BMPs on phosphorus removal.

| Moderator | Estimate (*lnRR*) | 95% CI | SE | $T$-statistic | df | $p$-value |
|---|---|---|---|---|---|---|
| Intercept | 0.40 | [-0.14,0.93] | 0.26 | 1.52 | 28 | 0.14 |
| log(Concentration) | 0.27 | [0.15,0.40] | 0.063 | 4.30 | 97 | <0.01 |

$I^2_{\text{total}}$=15.54, $I^2_{\text{study}}$=0, $I^2_{\text{effect}}$=15.54; $R^2_{\text{marginal}}$=0.32

### 3.3 Subsection 2

Frontiers requires figures to be submitted individually, in the same order as they are referred to in the manuscript. Figures will then be automatically embedded at the bottom of the submitted manuscript. Kindly ensure that each table and figure is mentioned in the text and in numerical order. Permission must be obtained for use of copyrighted material from other sources (including the web). Please note that it is compulsory to follow figure instructions. Figures which are not according to the guidelines will cause substantial delay during the production process.

## 4 DISCUSSION

Linking BMP performance to watershed scale improvements remains a challenge (Tomer and Locke, Melland et al, Meals). The succesful implementation and maintence of BMPs at watershed scales remains an area for more investigation (Liu et al., Lintern) Long-term trans-disciplinary monitoring projects not just on BMP performance but maintance, management, and social apsects may elucidate unknown factors influencing the ability to scale BMP reductions to watershed scale improvements (Lintern et al.)

Runoff volume was not incorporated into our models, largely attributed to lack of available data. However, BMP size is typically scaled to accommodate runoff volumes of a particular return period volume (1-yr or 2-yr flows for example) (Hoss et al.).

## DATA AVAILABILITY STATEMENT

The data and R code used in this study are deposited in Zenodo: TBD.

## DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## AUTHOR CONTRIBUTIONS

MS: Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing—original draft, Writing—review & editing. DK: Conceptualization, Formal analysis, Data collection, Writing—original draft, Writing—review & editing. JW: Data collection, Writing—review & editing. SJ: Data collection; Writing—review & editing.

## ACKNOWLEDGMENTS

Funding: This project was supported by a state nonpoint source grant from the Texas State Soil and Water Conservation Board.

## SUPPLEMENTAL DATA

Supplementary Material should be uploaded separately on submission, if there are Supplementary Figures, please include the caption in the same file as the figure. LaTeX Supplementary Material templates can be found in the Frontiers LaTeX folder

## REFERENCES

Agouridis, C. T., Workman, S. R., Warner, R. C., and Jennings, G. D. (2005). Livestock grazing management impacts on stream water quality: A review. *Journal of the American Water Resources Association* 41, 591–606. doi:10.1111/j.1752-1688.2005.tb03757.x

Allan, J. D. (2004). Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35, 257–284. doi:10.1146/annurev.ecolsys.35.120202.110122

Barrett, M. E. (2005). Performance comparison of structural stormwater best management practices. *Water Environment Research* 77, 78–86. doi:10.2175/106143005X41654

Barrett, M. E. (2008). Comparison of BMP performance using the International BMP Database. *Journal of Irrigation and Drainage Engineering* 134, 556–561. doi:10.1061/(ASCE)0733-9437(2008)134:5(556)

Benham, B., Zeckoski, R., and Yagow, G. (2008). Lessons learned from TMDL implementation case studies. *Water Practice* 2, 1–13. doi:10.2175/193317708X281370

Bernhardt, E. S., Band, L. E., Walsh, C. J., and Berke, P. E. (2008). Understanding, managing, and minimizing urban impacts on surface water nitrogen loading. *Annals of the New York Academy of Sciences* 1134, 61–96. doi:10.1196/annals.1439.014

Burnham, K. P., Anderson, D. R., and Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65, 23–35. doi:10.1007/s00265-010-1029-6

Buuren, S. V. and Groothuis-Oudshoorn, K. (2011). Mice: Multivariate Imputation by Chained Equations in *R. Journal of Statistical Software* 45. doi:10.18637/jss.v045.i03

Carey, R. O., Hochmuth, G. J., Martinez, C. J., Boyer, T. H., Dukes, M. D., Toor, G. S., et al. (2013). Evaluating nutrient impacts in urban watersheds: Challenges and research opportunities. *Environmental Pollution* 173, 138–149. doi:10.1016/j.envpol.2012.10.004

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., and Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications* 8, 559–568. doi:10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2

Cinar, O., Umbanhowar, J., Hoeksema, J. D., and Viechtbauer, W. (2021). Using information-theoretic approaches for model selection in meta-analysis. *Research Synthesis Methods* 12, 537–556. doi:10.1002/jrsm.1489

Clary, J., Leisenring, M., Poresky, A., Earles, A., and Jones, J. (2011). BMP performance analysis results for the International Stormwater BMP Database. In *World Environmental and Water Resources Congress 2011* (Palm Springs, California, United States: American Society of Civil Engineers), 441–449. doi:10.1061/41173(414)47

Cole, T. J. and Altman, D. G. (2017). Statistics Notes: What is a percentage difference? *BMJ* , j3663doi:10.1136/bmj.j3663

Collaboration for Environmental Evidence (2018). *Guidelines and Standards for Evidence Synthesis in Environmental Management*. Tech. Rep. Version 5.0

EPA (2017). *National Water Quality Inventory: Report to Congress*. Tech. Rep. EPA 841-R-16-011, U.S. Environmental Protection Agency, Washington, D.C.

Freeman, L. A., Corbett, D. R., Fitzgerald, A. M., Lemley, D. A., Quigg, A., and Steppe, C. N. (2019). Impacts of urbanization and development on estuarine ecosystems and water quality. *Estuaries and Coasts* 42, 1821–1838. doi:10.1007/s12237-019-00597-z

Fu, B., Merritt, W. S., Croke, B. F., Weber, T. R., and Jakeman, A. J. (2019). A review of catchment-scale water quality and erosion models and a synthesis of future prospects. *Environmental Modelling & Software* 114, 75–97. doi:10.1016/j.envsoft.2018.12.008

Grudzinski, B., Fritz, K., and Dodds, W. (2020). Does riparian fencing protect stream water quality in cattle-grazed lands? *Environmental Management* 66, 121–135. doi:10.1007/s00267-020-01297-2

Hedges, L. V., Gurevitch, J., and Curtis, P. S. (1999). The meta-analysis of response ratios in experimental ecology. *Ecology* 80, 1150–1156. doi:10.1890/0012-9658(1999)080[1150:TMAORR]2.0.CO;2

Horvath, I. R., Zhang, K., Mayer, B. K., and Parolari, A. J. (2023). Effects of regional climate and BMP type on stormwater nutrient concentrations in BMPs: A meta-analysis. *Environmental Science & Technology* 57, 5079–5088. doi:10.1021/acs.est.2c05942

Kambach, S., Bruelheide, H., Gerstner, K., Gurevitch, J., Beckmann, M., and Seppelt, R. (2020). Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis. *Ecology and Evolution* 10, 11699–11712. doi:10.1002/ece3.6806

Koch, B. J., Febria, C. M., Gevrey, M., Wainger, L. A., and Palmer, M. A. (2014). Nitrogen removal by stormwater management structures: A data synthesis. *JAWRA Journal of the American Water Resources Association* 50, 1594–1607. doi:10.1111/jawr.12223

Kroger, R., Perez, M., Walker, S., and Sharpley, A. (2012). Review of best management practice reduction efficiencies in the Lower Mississippi Alluvial Valley. *Journal of Soil and Water Conservation* 67, 556–563. doi:10.2489/jswc.67.6.556

Lintern, A., McPhillips, L., Winfrey, B., Duncan, J., and Grady, C. (2020). Best management practices for diffuse nutrient pollution: Wicked problems across urban and agricultural watersheds. *Environmental Science & Technology* 54, 9159–9174. doi:10.1021/acs.est.9b07511

Liu, Y., Ahiablame, L. M., Bralts, V. F., and Engel, B. A. (2015). Enhancing a rainfall-runoff model to assess the impacts of BMPs and LID practices on storm runoff. *Journal of Environmental Management* 147, 12–23. doi:10.1016/j.jenvman.2014.09.005

Liu, Y., Engel, B. A., Flanagan, D. C., Gitau, M. W., McMillan, S. K., and Chaubey, I. (2017). A review on effectiveness of best management practices in improving hydrology and water quality: Needs and opportunities. *Science of The Total Environment* 601–602, 580–593. doi:10.1016/j.scitotenv.2017.05.212

Meals, D. W., Dressing, S. A., and Davenport, T. E. (2010). Lag time in water quality response to best management practices: A review. *Journal of Environment Quality* 39, 85. doi:10.2134/jeq2009.0108

Nakagawa, S. and Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology* 26, 1253–1274. doi:10.1007/s10682-012-9555-5

Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtaining $R^2$ from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4, 133–142. doi:10.1111/j.2041-210x.2012.00261.x

Nakagawa, S., Yang, Y., Macartney, E. L., Spake, R., and Lagisz, M. (2023). Quantitative evidence synthesis: A practical guide on meta-analysis, meta-regression, and publication bias tests for environmental sciences. *Environmental Evidence* 12, 8. doi:10.1186/s13750-023-00301-6

National Research Council (2001). *Assessing the TMDL Approach to Water Quality Management* (Washington, D.C.: National Academies Press). doi:10.17226/10146

Nuzzo, R. (2018). Percent differences: Another look. *PM&R* 10, 661–664. doi:10.1016/j.pmrj.2018.05.003

Osenberg, C. W., Sarnelle, O., and Cooper, S. D. (1997). Effect size in ecological experiments: The application of biological models in meta-analysis. *The American Naturalist* 150, 798–812. doi:10.1086/286095

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria

Rohatgi, A. (2022). *WebPlotDigitizer*. Pacifica, California

Schramm, M., Gitter, A., and Gregory, L. (2022). Total Maximum Daily Loads and *Escherichia coli* trends in Texas freshwater streams. *Journal of Contemporary Water Research & Education* 176, 36–49. doi:10.1111/j.1936-704X.2022.3374.x

Simpson, D. T. and Weammert, S. (2009). *Developing Best Management Practice Definitions and Effectiveness Estimates for Nitrogen, Phosphorus and Sediment in the Chesapeake Bay Watershed*. Tech. rep., University of Maryland Mid-Atlantic Water Program

Tasdighi, A., Arabi, M., Harmel, D., and Line, D. (2018). A Bayesian total uncertainty analysis framework for assessment of management practices using watershed models. *Environmental Modelling & Software* 108, 240–252. doi:10.1016/j.envsoft.2018.08.006

Ullrich, A. and Volk, M. (2009). Application of the Soil and Water Assessment Tool (SWAT) to predict the impact of alternative management practices on water quality and quantity. *Agricultural Water Management* 96, 1207–1217. doi:10.1016/j.agwat.2009.03.010

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 36. doi:10.18637/jss.v036.i03

## FIGURE CAPTIONS

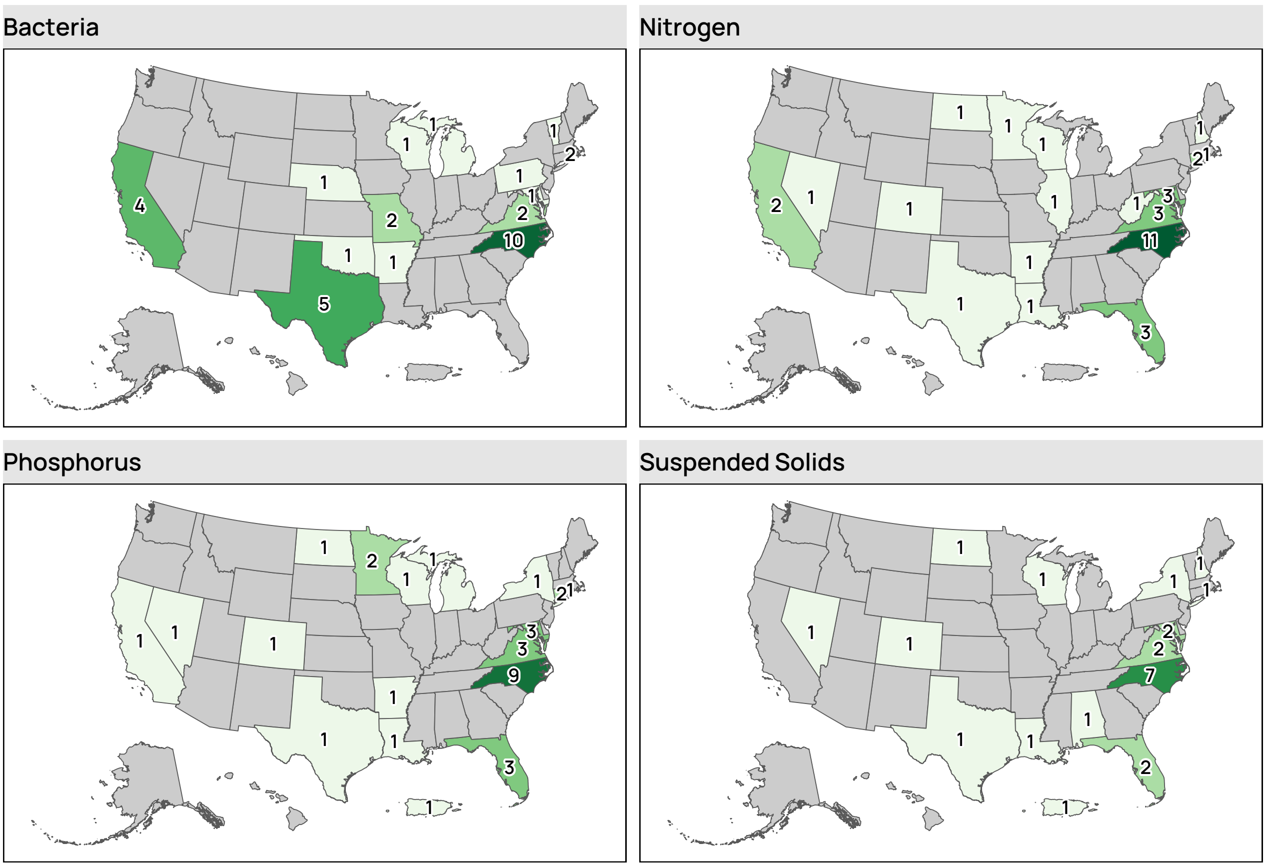


**Figure 1.** Distribution of studies identified in the systematic review by state and parameter.

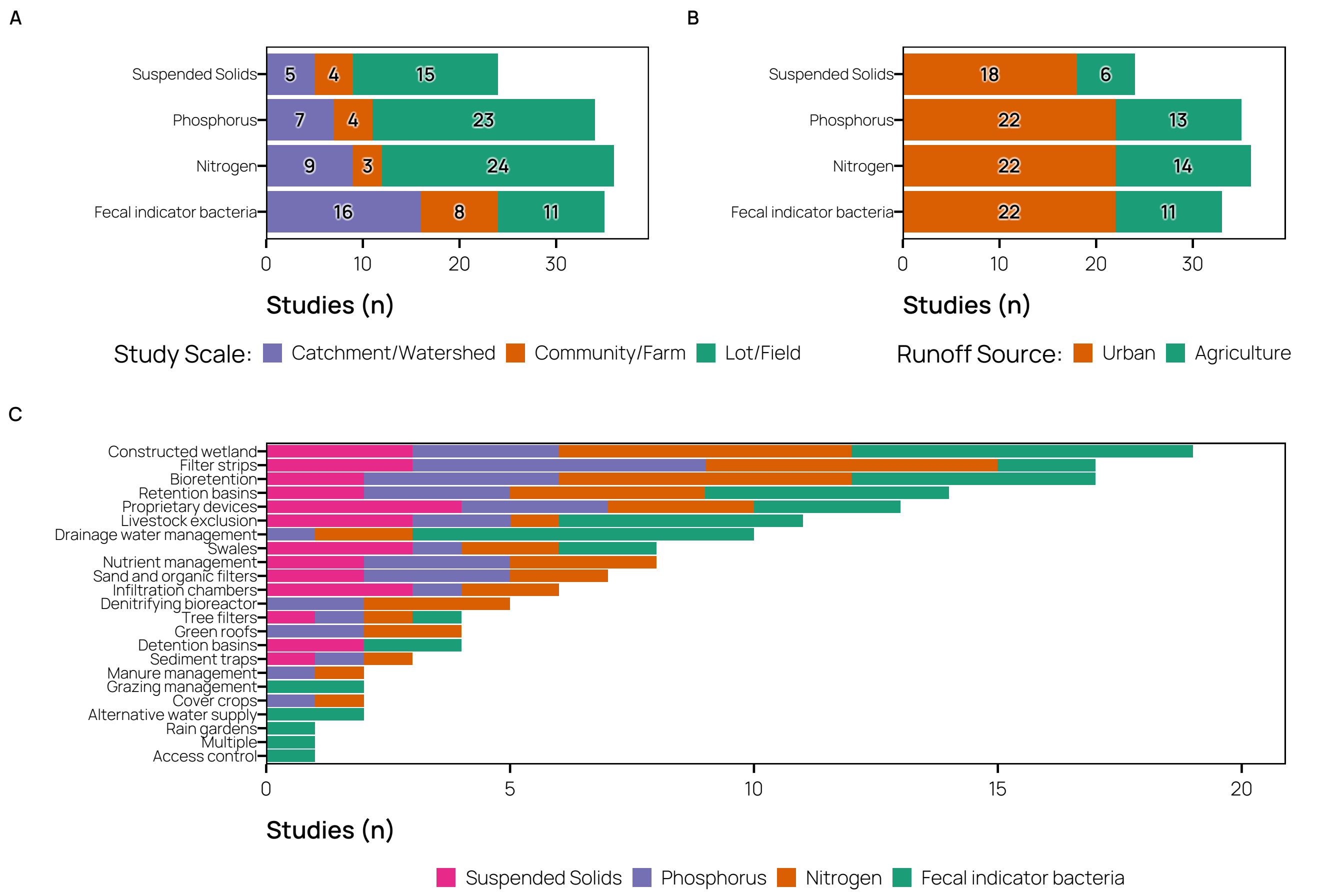


**Figure 2.** Summary of (A) study scale, (B) dominant runoff source, and (C) BMPs identified in the systematic review.