Journal Pre-proofs

Full length article

Recreational exposure to polluted open water and infection: A systematic review and *meta*-analysis protocol

Matt Lloyd Jones, Anne Clare Frances Leonard, Alison Bethel, Emma Lamb, William H. Gaze, Tim Taylor, Andrew C. Singer, Obioha C. Ukoumunne, Ruth Garside

PII: S0160-4120(25)00122-9
DOI: https://doi.org/10.1016/j.envint.2025.109371
Reference: EI 109371

To appear in: *Environment International*

Received Date: 23 July 2024
Revised Date: 7 February 2025
Accepted Date: 10 March 2025

Please cite this article as: M.L. Jones, A.C. Frances Leonard, A. Bethel, E. Lamb, W.H. Gaze, T. Taylor, A.C. Singer, O.C. Ukoumunne, R. Garside, Recreational exposure to polluted open water and infection: A systematic review and *meta*-analysis protocol, *Environment International* (2025), doi: https://doi.org/10.1016/j.envint.2025.109371

Recreational exposure to polluted open water and infection: A systematic review and meta-analysis protocol

Matt Lloyd Jones†[1], Anne Clare Frances Leonard†[2], Alison Bethel†‡[3], Emma Lamb†[4], William H. Gaze†[5], Tim Taylor†[6], Andrew C. Singer⊥[7], Obioha C. Ukoumunne‡[8], Ruth Garside†[9]

† European Centre for Environment and Human Health, Peter Lanyon Building, University of Exeter Penryn,TR10 8RD

‡ NIHR CLAHRC South West Peninsula, University of Exeter Medical School, South Cloisters, St Luke's Campus, Exeter, Devon, United Kingdom

ǂ NIHR ARC South West Peninsula (PenARC), University of Exeter Medical School, Exeter, UK

⊥ UK Centre for Ecology & Hydrology, Wallingford, UK.

Contact addresses:

1. M.l.jones@exeter.ac.uk (Corresponding)
2. Anne.Leonard@exeter.ac.uk
3. A.Bethel@exeter.ac.uk
4. E.Lamb@exeter.ac.uk
5. W.H.Gaze@exeter.ac.uk
6. Timothy.J.Taylor@exeter.ac.uk
7. acsi@ceh.ac.uk
8. O.C.Ukoumunne@exeter.ac.uk
9. R.Garside@exeter.ac.uk

Abstract

Background: Open water recreation (e.g. swimming, surfing) is growing in popularity alongside concerns about contracting infections as a result of wastewater (including sewage) and runoff pollution in seas, rivers, lakes, and other bodies of open water. Previous systematic reviews have found evidence for a positive association between exposure to open water and infection. However, these syntheses focus on comparisons of recreational water users and non-recreational water users, and make concessions on key stages of the systematic review process. This limits their ability to summarise the evidence for an effect of exposure to pollution, specifically.

Methods: We present a peer-reviewed protocol for a systematic review and meta-analysis of exposure to wastewater and runoff pollution and infection in recreational open water users in the Global North. Eligible studies must contain at least two groups of recreational water users known or suspected to have been exposed to distinct levels of pollution, with some estimate of cases of infection in each group. These studies will be obtained via searches of bibliographic databases (MEDLINE, Web of Science Core Collection, Environment Complete, and Global Health), grey literature sources, and supplementary search methods. Risk of bias in these studies will be assessed using Cochrane's ROBINS-E and RoB 2 tools. Studies' results will be qualitatively and quantitatively synthesised, following and reporting to contemporary standards and guidelines (e.g. PRISMA, SWiM). The results of the review will be summarised with a GRADE certainty assessment of the evidence for different types of infections, presented in a Summary of Findings table.

Rationale

Background

Open water bathing activities such as open water swimming and surfing are increasingly popular, not least for their physical and mental health benefits (The Big Blue Swim, 2020; Lemmin-Woolfrey, 2021; Carroll, 2021; The Nursery Research and Planning, 2021, 2022; Sasse, 2024; Britton et al., 2020). Nonetheless, their rising popularity is occurring alongside growing concerns about anthropogenic pollution of natural waters. In the United Kingdom, the issue is particularly high on the public and policy agenda (Environment Agency, 2024a, 2024b, 2023a; Hammond et al., 2021; Laville et al., 2024; Mathers, 2024; National Engineering Policy Centre Working Group, 2024). Concern is largely due to the consistently high volumes of raw sewage being discharged into UK seas, rivers and lakes via combined sewage overflows (3.6 million hours' recorded in 2023; Burnett and Bolton, 2025) - but also growing concerns about agricultural runoff pollution in iconic bathing locations such as the Wye and Windermere (DEFRA, 2024; Environment Agency, 2024c**)**. Beyond the UK, even in countries without combined sewerage and/or which have invested more heavily in 'cleaning up' bathing waters, there is also a risk that treated wastewater and runoff pollution is still releasing dangerously high numbers of waste-derived microorganisms into natural waters (National Engineering Policy Centre Working Group, 2024). This was recently highlighted by discussions around the risk that swimming in the Seine as part of the 2024 Paris Olympics posed to athletes, despite substantial efforts to improve water quality before the event (Park, 2024; Reynolds et al., 2024; Tien, 2024).

There is thus renewed public and policy demand for more information about whether wastewater and runoff pollution pose a colonisation and infection risk to open water recreational water users – demand which is only likely to grow as climate change alters rainfall patterns in ways that put additional pressure on water and waste management infrastructure (Gogien et al., 2023; Hensyl et al., 2024; OFWAT, 2008). To some extent, this demand has already been met with a relatively long history of scientific studies in this area and several syntheses of their evidence (Tables 1 & 2; Supplementary Information 1). These syntheses include a systematic review conducted by members of our own team between 2013 and 2018 (Leonard et al., 2018a), which synthesised evidence from non-randomised and randomised studies available at the time (last searches conducted 2015). The meta-analyses emerging from this review found associations between recreational exposure to seawater and experiencing symptoms of any infection (odds ratio (OR) = 1.86, 95% confidence interval (CI): 1.31 to 2.64, P = 0.001), ear infections specifically (OR = 2.05, 95% CI: 1.49 to 2.82, P < 0.001), and gastrointestinal infections specifically (OR = 1.29, 95% CI: 1.12 to 1.49, P < 0.001). A more recent systematic review (Russo et al., 2020) also quantified the association between recreational exposure to seawater and freshwater and indicators of gastrointestinal infection, and found an even stronger association (particularly in freshwater).

These syntheses support the notion that bathing in open water is associated with infection, and in particular have highlighted that the strongest associations are likely to be found in high-contact activities such as swimming and surfing (Russo et al., 2020). Nonetheless, as well as being at least 5 years out of date, these syntheses have focused on studies that compare groups not exposed to open water with those exposed to it (which are the most numerous in the literature) and (probably due to the resulting scale of the synthesis task) make concessions on several key stages of the systematic review process. Whilst understandable, this limits the extent to which they can be used to assess the evidence for a relationship specifically between exposure to wastewater and runoff pollution and infections. It also limits their scope to explore which sub-exposures/outcomes are most strongly associated with wastewater and run-off pollution, and the extent to which any relationships found are likely to causative. Here, we propose a more focussed but more detailed systematic review of the evidence relating to the role of wastewater and runoff pollution in infection in recreational open water users specifically, incorporating the latest evidence.

Motivation for the systematic review

Shifting towards a more direct assessment of the role of wastewater and runoff pollution

Both Leonard et al., (2018a) and Russo et al., (2020) paid some attention to studies investigating the effects of exposure to wastewater and runoff pollution specifically (hereafter 'pollution exposures'). However, in both reviews this was obscured by the synthesis of more-studied exposures to open water *per se* (hereafter 'water exposures'). Leonard et al., (2018a) included pollution exposures in its Study Characteristics table, but the remainder of the synthesis focussed exclusively on (sea)water exposures. Russo et al., (2020) meanwhile, pooled pollution exposures together with water exposures, blurring the distinction between the two in the synthesis results. Both choices are understandable, given that, compared to those working in clinical fields, synthesists working on environmental health and toxicology-type questions are often compelled make the most of a more scant and heterogeneous evidence base (Wikoff et al., 2020). Nonetheless, there are several reasons why a dedicated synthesis of the pollution exposures may better clarify the role of wastewater and run-off pollution specifically.

Primarily, this is because estimating the effect of water exposure does not directly address the effect of pollution exposure (i.e. inferences about pollution exposure are confounded by water exposure). Indeed, Leonard et al., (2018a) found the largest association between exposure to seawater and reporting symptoms of ear infections, which could be for reasons other than exposure to pollution. For example, surfer's ear is an exostosis associated with open water bathing (including surfing) that may present similar symptoms to an infection or make outdoor water users more vulnerable to infection (Kroon et al., 2002). However, its ultimate cause is exposure to cold wind/water rather than infectious microorganisms. Recreational water users are also more susceptible to respiratory and skin conditions that do not have microbial causes but which may present similar symptoms (Smith et al., 2018;

Freiman et al., 2004). These examples highlight why water exposures may not always be a reliable proxy for pollution exposures - especially when relying on self-reported symptoms.

This issue can be mitigated by the synthesising additional evidence from included studies which further supports the presence of infectious agents at the suspected site of exposure and/or infection (and ideally a dose-response relationship between the two). However, compared to pollution exposures, water exposures are still arguably more vulnerable to confounding at the population level and lack of blinding. On the former, recreational water users are a demographically different group to non-recreational water users (Outdoor Swimmer, 2021; Sport England, 2024; Surf Industry Members Association, 2024), confounding the exposure in ways that are hard to fully control for (particularly in the non-randomised designs that dominate this field; Kay et al., 1994). This may warrant them being considered a distinct population (as we do in this review; see Methods: Eligibility Criteria). Regarding lack of blinding, like many exposures in environmental health (Allen et al., 2015), water exposure is inherently hard to blind; at least to participants, it is obvious who has been in the water (even in randomised studies). This may be particularly problematic in this field, since it is known that 'risk perception bias' – a phenomenon in which participants' likelihood of reporting symptoms is related to the risk they perceive to be associated with bathing – can cause those knowingly exposed to water to report inflated rates of infection (Fleisher and Kay, 2006). Pollution exposures are probably less vulnerable to this, since the degree of pollution is not immediately obvious unless pollution is extreme (e.g. obvious visual or olfactory signs), or participants and researchers have checked water quality information prior to exposure (though systematic reviewers should still try to validate and account for this where necessary).

We therefore believe a dedicated synthesis of pollution exposures could provide a fuller understanding of the evidence for wastewater and runoff pollution driving infection, and believe there are sufficient studies of such exposures to justify a review. Leonard et al., (2018a) listed (but did not synthesise) 6 studies including pollution exposures conducted in the Global North in its Study Characteristics table. We have further identified at least 2 studies that may have been omitted from Leonard et al., (2018a)'s list, and at least 2 new studies that could be included in a dedicated synthesis of the effect of pollution exposure (Supplementary Information 2). We have not yet performed a dedicated search for further new studies (as this is partly the point of the review), but we anticipate this number to increase once the new search strategy has been executed. It has previously been estimated that the median time for systematic reviews in the medical sciences to become out of date (defined as a change in statistical significance or relative change in the magnitude of effects involving key outcomes of at least 50%) is 5.5 years (Shojania et al., 2007). Five years have passed since the publication of Russo et al., (2020) and seven since the publication of Leonard et al., (2018a), with the searches underpinning these reviews even more out of date. Especially with an improved search strategy, we are therefore confident that a new review will identify sufficient studies to synthesise, and that there is ample opportunity to gain new insights by synthesising it with more methodological rigour.

Another motivation for conducting a new systematic review in this area is that syntheses conducted to date make important concessions on key stages of a systematic review, particularly in the pre-registration and quality appraisal stages (Table 1, Table 2). Whilst such syntheses still provide many useful insights that inform the development of this review, this nonetheless limits their ability to provide an unbiased assessment of the strength of the evidence base (as well as the strength of effects reported within it).

Regarding pre-registration, several reviews in this field lack a pre-registered protocol altogether (Mannocci et al., 2016; Russo et al., 2020; Wade et al., 2003; Yau et al., 2009). Although pre-registration does not in itself guarantee the quality of a review (though there is some emerging evidence that it increases the chances of a better quality review; Mei *et al.*, 2022), it does at least allow the reader to distinguish between pre-specified and post-hoc synthesis choices that may have an important bearing on the results. A relevant example in this regard is Russo et al., (2020)'s subgroup analysis by water quality, in which the water exposures synthesised are categorised "good," "intermediate," or "poor" (deduced from how it was reported in the studies). This decision has an important bearing on the conclusion that reported effects of water exposure are likely due to pollution exposure. However, given the lack of pre-registration, it is not possible to know whether the choice of water quality categories could have been implicitly influenced by the synthesists' knowledge of the outcome data (especially as it is quite a subjective categorisation). To minimise such risks of bias, pre-registration has become the norm in systematic review, and systematic review guidelines and journals strongly encourage, and in some cases explicitly require it. Accordingly, protocols for the reviews of (Leonard et al., 2018a) and Adhikary et al., (2022) were pre-registered, though not peer-reviewed. Building on this, this protocol is first among the systematic reviews identified to have also been peer-reviewed prior to its pre-registration/publication.

There are also significant opportunities to improve upon the risk of bias and certainty assessment stages of systematic reviews in this area. These are key stages of a systematic review that enable a systematic assessment of the strengths and weaknesses of the current evidence base, and meta-analytical results to be fully contextualised (Izcovich et al., 2022). Regarding risk of bias, some reviews did not conduct risk of bias assessment at all (Wade et al., 2003; Yau et al., 2009) and those that did conducted it using tools that are now considered inadequate (Table 1, Table 2). For example, our own review (Leonard et al., 2018a) used the Critical Appraisal Skills Programme tool and others (Mannocci et al., 2016; Adhikary et al., 2022) used the Newcastle-Ottawa Scale – both of which have been criticised as being under-sensitive (Hannes et al., 2010; Stang, 2010; Stang et al., 2018). To be sure, at the time some of these reviews were conducted, Cochrane's ROBINS-I (Sterne et al., 2016) and ROBINS-E (Higgins et al., 2024) tools - which are now considered the gold standard for assessing observational studies that dominate this field – were not available. This is probably why such reviews focussed on more generalist tools that could be applied to both randomised and non-randomised studies. Nonetheless, there is still an opportunity to improve synthesis methods

by conduct risk of bias assessment with the best currently available tools. Furthermore, none of the previous reviews conducted a certainty assessment (e.g. GRADE; Schünemann et al., 2013) bringing together risk of bias assessments with other evidence from the review to systematically summarise the confidence in the cumulative evidence. We consider this an especially important omission when synthesising these environmental health studies, in which it is not always practical to establish either the exposure or outcome with great certainty (Morgan et al., 2018). A certainty assessment would greatly help contextualise the key meta-analytical results for practitioners, summarising the extent to which the least/most certain studies contribute to estimates of association.

We also think that there are opportunities to improve upon other stages of the systematic review process and their reporting such as the framing of the objectives, search strategy, and meta-analyses (Table 1; Table 2). Granted, the use of systematic review and meta-analysis in environmental health is still relatively nascent, and some standards have been raised in the field after many of these reviews were initiated (Whaley et al., 2016; Haddaway et al., 2018; Whaley and Roth, 2022; Nakagawa et al., 2023). Nonetheless, many of these standards have at least been long-established in clinical research from which systematic review process is derived (Moher et al., 1999, 2009; Shamseer et al., 2015), and we think there is a clear opportunity of a more focussed review that better adheres to them.

Table 1: Similar systematic reviews published since (and including)

Study	Leonard et al., (2018a)	Russo et al., (2020)	Adhikary et al. (2022)
Pre-registered protocol	PROSPERO CRD42013005307)	Not pre-registered	PROSPERO CRD42019145265
Objective	*"In this systematic review, the aim was to gather, appraise and synthesize the evidence to answer the following research questions.* *i. Do bathers have an increased risk of experiencing symptoms of infections following recreational use of coastal water, compared with non-bathers? ii. Does the level of increased risk depend on the nature of exposure to the water?"*	*"The objectives of this study are to: (1) assess and summarize the scientific literature on the risk of illness associated with different types of recreational activities and different levels of water contact during recreation; (2) quantitatively estimate the pooled risk of illness associated with different categories of recreational activities and levels of water contact; and (3) evaluate risk of illness across activity and water contact categories to better understand illness risk associated with different types of recreation in ambient surface waters."*	"*To examine the health risks associated with microbial exposure from the recreational use of untreated freshwater bodies and assess whether the risks differed by the type of contact (primary, secondary, no contact) and waterbody."*
Population	Healthy people living in OECD countries	Any	Any
Exposure	Recreational exposure to seawater* (full qualitative and quantitative synthesis); recreational exposure to more polluted seawater* (study characteristics table only)	Recreational exposure to seawater or freshwater	Recreational exposure to freshwater
Comparator	No recreational exposure to seawater (full qualitative and quantitative synthesis); recreational exposure to least polluted seawater (study characteristics table only)	No recreational exposure to seawater[?]; recreational exposure to least polluted seawater[?]	No recreational exposure to freshwater
Outcomes	Self-reported symptoms or laboratory infection (any, ear, gastrointestinal, eye, respiratory, skin, urinary, and other in narrative synthesis; any, ear, and gastrointestinal in meta-analyses)	Self-reported symptoms or laboratory infection; laboratory-confirmed infection (ear, gastrointestinal, eye, respiratory, skin, urinary, and other in narrative synthesis; gastrointestinal and respiratory in meta-analyses)	Self-reported symptoms or laboratory infection (gastrointestinal, ear, eye, and skin in narrative synthesis; no meta-analyses)
Study designs	Non-randomised; randomised	Non-randomised; randomised	Non-randomised; randomised
Bibliographic databases searched	BIOSIS; EMBASE; Environment Complete; GreenFILE; MEDLINE; Web of Science	PubMed; Web of Science; TOXLINE	CAB Abstracts; PubMed; Scopus; Web of Science
Search dates	July 2013 (updated June 2015)	Not reported	August 2019 (updated June 2021)
Number of included studies	40 (19 in meta-analyses)	92 in meta-analyses (no narrative synthesis of studies not included in meta-analyses)	35 (no meta-analyses)
Risk of bias assessment method	Critical Appraisal Skills Programme (CASP) tool	Ad-hoc, hybrid tool developed with inspiration from the NTP OHAT tool, Agency for Healthcare Research and Quality, Cochrane Handbook, and CLARITY Group	Newcastle-Ottawa Quality Assessment Scale (NOS) for non-randomised studies†; Newcastle-Ottawa Quality Assessment Scale (NOS) adapted for randomised controlled trials
Narrative synthesis	Study characteristics table; narrative description of study characteristics	Study characteristics infographic; Study characteristics table; narrative description of study characteristics	Study characteristics table; narrative description of study characteristics; vote counting based on statistical significant results across studies[?]
Quantitative synthesis (meta-analyses)	Multiple independent random effects models (10) pooling odds ratios for no seawater exposure/seawater exposure comparisons in each of 10 symptom categories	Multiple mixed-effects/meta-regression models (unclear how many) combining odds ratios for no seawater exposure/seawater exposure and least pollution exposure/more pollution exposure comparisons into pooled effects for many different categories of moderators	None
Certainty assessment	None	None	None
Main conclusions	Higher risk of experiencing symptoms of any illness, ear infections (highest risk), and gastrointestinal infections in those exposed to seawater versus those not exposed to seawater.	Higher risk of experiencing symptoms of gastrointestinal infection and respiratory infection in those exposed to seawater versus those with minimal or no exposure to seawater.	Most studies reported higher risk of experiencing symptoms of infection in those exposed to freshwater versus those with minimal or no exposure to seawater.

*Freshwater exposure was originally intended to be included in Leonard et al., (2018a) and as such, freshwater terms were included in the search strategy. However, as the authors note *"Given the scale of the evidence base and resource constraints, the freshwater papers were not fully data-extracted or quality-appraised, and have not been included in this review."*.

[?]Inferred as was not clearly reported in the study

†The Newcastle-Ottawa Scale (NOS) has been criticised for being untransparent in its methods, omitting key elements of bias, and including potentially invalid items (Stang, 2010; Stang et al., 2018)

Table 2: Similar systematic reviews published before Leonard et al., (2018a)

Study	Wade et al., (2003)	Yau et al. (2009)	Mannocci et al. (2016)
Pre-registered protocol	Not pre-registered	Not pre-registered	Not pre-registered
Objective	*"Our primary goal in this systematic review was to evaluate the evidence linking specific microbial indicators of recreational water quality to specific health outcomes under nonoutbreak conditions. Secondary goals were to identify and describe critical study design issues, to quantify and evaluate sources of heterogeneity among the studies, and to evaluate the potential for health effects at or below the current suggested regulatory standards."*	*"The primary goal of this investigation is to quantify the association between microbial indicators used to monitor recreational water quality and skin-related outcomes in non-outbreak conditions in both marine and freshwater settings."*	*"the aim of the present study is to systematically review and meta-analyze the association between swimming in recreational water and the occurrence of respiratory illness."*
Population	Swimmers	Any	Any (other than professional swimmers)
Exposure	Recreational exposure to seawater; recreational exposure to more polluted seawater	Recreational exposure to seawater or freshwater	Recreational exposure to seawater; [?]
Comparator	No recreational exposure to seawater[?]; recreational exposure to least polluted seawater[?]	No recreational exposure to seawater[?]; recreational exposure to least polluted seawater[?]	No recreational exposure to freshwater[?]
Outcomes	Self-reported symptoms or laboratory infection (gastrointestinal in narrative synthesis; gastrointestinal in meta-analysis)	Self-reported symptoms or laboratory infection (skin in narrative synthesis; skin in meta-analyses)	Self-reported symptoms or laboratory infection (respiratory in narrative synthesis; respiratory in meta-analyses)
Study designs	Non-randomised; randomised	Non-randomised; randomised	Non-randomised; randomised
Bibliographic databases searched	BIOSIS; EMBASE; MEDLINE, OLDMEDLINE; ProQuest	BIOSIS; EMBASE; PubMed; Web of Science; ProQuest	PubMed; Scopus
Search dates	Not reported	August 2008[?]	February 2015[?]
Number of included studies	27 in meta-analyses (no narrative synthesis of studies not included in meta-analyses)	20 in meta-analyses (no narrative synthesis of studies not included in meta-analyses)	14 in meta-analyses (no narrative synthesis of studies not included in meta-analyses)
Risk of bias assessment method	None	None	Newcastle-Ottawa Quality Assessment Scale (NOS) for non-randomised studies; Jadad scale for randomised studies†
Narrative synthesis	Study characteristics table; narrative description of study characteristics; summarising risk ratios; vote counting based on correlation coefficients and p-values	Study characteristics table; narrative description of study characteristics	Study characteristics table; narrative description of study characteristics
Quantitative synthesis (meta-analyses)	Multiple independent random effects models (unclear how many) combining risk ratios for no seawater exposure/seawater exposure and least pollution exposure/more pollution exposure comparisons into pooled effects for many different categories of moderators; meta-regression of risk ratios against indicator bacteria density	Multiple independent fixed and random effects models (unclear how many) combining risk ratios for no seawater exposure/seawater exposure and least pollution exposure/more pollution exposure comparisons into pooled effects for many different categories of moderators for many different categories of moderators	Random effect model pooling risk ratios for no seawater exposure/seawater exposure
Certainty assessment	None	None	None
Main conclusions	Higher risk of experiencing symptoms of gastrointestinal infection in those exposed to seawater versus those not exposed to seawater. Risk correlates with microbial indicators of pollution.	Higher risk of experiencing symptoms of skin infection in those exposed to polluted seawater versus those not exposed to seawater	Higher risk of experiencing symptoms of respiratory infection in those exposed to seawater versus those not exposed to seawater.

[?] Inferred as was not clearly reported in the study

†The Newcastle-Ottawa Scale (NOS) has been criticised for being untransparent in its methods, omitting key elements of bias, and including potentially invalid items (Stang, 2010; Stang et al., 2018)

Table 3: PECO-structured review objectives for the systematic review and meta-analysis

Category	**Criteria**
Population	Recreational open water users in Global North locations
Exposure	Open water judged as being more affected by wastewater and runoff pollution
Comparator	Open water judged as being less affected by wastewater and runoff pollution
Outcomes	Estimated cases of different types of infection (ear, eye, gastrointestinal, skin, respiratory tract, and urinary tract infections) by microorganisms. The term 'estimated cases' covers cases estimated from self-reported symptoms, diagnostic, or microbiological confirmation.

Objectives

In line with its motivations, the main objective of the systematic review and meta-analysis we are proposing is to answer the following PECO-structured research question (Morgan et al., 2018):

Q1. In studies of recreational open water users in Global North locations (P), what is the association between exposure to open water judged as being more affected by wastewater and run-off pollution (E) versus exposure to open water judged as being the least affected by pollution (C) on estimated cases of infection (including colonisation; O)?

Capitalising on our diverse experience (Table S1) we intend to narratively and quantitatively synthesise the results of studies sharing these objectives (Table 3). In order to do this, we will carry out up-to-date searches for relevant studies, summarise study characteristics and risk of bias, before carrying out quantitative synthesis (meta-analysis). The quantitative synthesis will primarily provide estimates of the association between recreational exposure to open water judged as being polluted and estimated cases of infection of the ear, and the gastrointestinal tract (which we previously assessed for non-seawater users versus seawater users) – as well as of the eye, skin, respiratory tract, urinary tract and other types. This will be followed by additional analyses and certainty assessments that seek to carefully consider the extent to which the results of these primary analyses are affected by varying definitions of 'polluted water' in the literature. Finally, all of the information from the systematic review and meta-analysis will be drawn together into a summary of findings table including certainty assessment. Ultimately, our aim is to provide researchers, policymakers and other stakeholders with the most robust assessment to date of the evidence for an association between exposure to wastewater and run-off pollution, and infection by microorganisms.

Methods

Eligibility criteria

Eligibility criteria are closely based on the PECO-structured review objectives. These criteria will be used during the screening of titles and abstracts, and full texts of reports (see Selection process) to identify primary studies with non-randomised designs (e.g. cohort studies, cross-sectional studies) and randomised designs (e.g. randomised controlled trials, cluster-randomised controlled trials). Studies will be included if they contain study groups meeting the following characteristics.

Population

Recreational open water users: People of any age who have taken part in immersive water activities in Global North locations in the last 1 month. 'Immersive water activities' includes activities like swimming and surfing, where body immersion is necessary and head immersion likely. 'Global North locations' means bathing and recreational waters(not that recreational water users are necessarily from these destinations) in Global North countries as defined by UNCTAD (i.e. Europe - Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom, Georgia, Ukraine, Moldova; North America – Canada, United States; Japan; South Korea; Australia; New Zealand; Israel). Global South and overseas territories of Global North member countries are excluded (e.g. Puerto Rico, British Virgin Islands). Nonetheless, reviewers should tag studies from Global South and other countries during the abstract screening phase.

Justification: Recreational water users are demographically different to non-recreational water users, and restricting the definition to activities where both body and head immersion are unavoidable reduces heterogeneity arising from lower contact activities, which have been shown to be associated with a lower risk of illness in previous reviews (Leonard et al., 2018a; Russo et al., 2020)*. A recall period of 1 month is used because of the varying time-to-onset of different infections and varying recall periods within and between studies. The population was limited to recreational water users in Global North destinations to reduce heterogeneity that needs to be dealt with in the synthesis. Bathing water quality is likely to be established with less certainty in studies of non-Global North destinations (due to less-funded studies and lack of secondary data sources) and of lower quality than in Global North destinations (Malik et al., 2014). We used the UNCTAD definition of Global North countries as of February 2025 (UNCTAD, 2024, 2023, 2022), because it provides an independent, economically-based categorisation upon which base our eligibility criteria.* Global South and other countries will be tagged during the screening phase to create a collection of these studies that will interested readers and future synthesists to pursue them further, in order to compensate somewhat for the focus on the Global North.

Exposure

Higher pollution: At least one group of recreational water users known or suspected of having been exposed to open water more affected by untreated/treated wastewater (including sewage) and runoff pollution than the Comparator group. Where exposure is not established with certainty, suspicion of exposure may come from anecdotal or circumstantial information (e.g. proximity to likely point sources), bathing water measures/status from a source external to the study (e.g. environment agency), or direct measures of indicator organisms. Assessors should not include exposures where pollution is reported to be associated with algae/algal blooms and the outcome is not confirmed to be an infection (i.e. based only on symptoms).

Justification: Wastewater and runoff pollution are the main forms of pollution that are on the public and policy agenda regarding human infection, given their likelihood of carrying human/animal faecal microorganisms and/or amplifying harmful microorganisms in the environment. In most studies, we anticipate that participants' exposure to pollution will not be established with absolute certainty, and will instead come from indirect signs that lead the investigators to believe that exposure has occurred. Hence, we are relatively astringent about the criteria used to establish pollution exposure - especially as we intend to compensate for this by directly discussing the potential effect of the directness of the pollution measure through our narrative and quantitative syntheses (see Additional analyses). Multiple groups which have higher exposure than the comparator may be present in studies (e.g. multiple polluted bathing destinations or ordinal groups of a pollution exposure). These are eligible to be included, with shared control non-independence will be accounted for in the meta-analytical models used in the quantitative analysis. Regarding the exclusion of algal-mediated exposures without confirmation of infection, this is because symptoms may be attributable to toxins in such cases, and we are specifically interested in infection. Of course, there is a risk that poorly measured exposures/outcomes will be included even though they are influenced by algae/toxin-mediated health effects. However, this will be accounted for in risk of bias/certainty of evidence assessments in the same way as outcomes only based on symptoms are assigned less certainty.

Comparison

Lowest pollution: A group of recreational water users known or suspected of having been exposed to open water that is the least affected by wastewater (including sewage) and runoff pollution. Where exposure is not established with certainty, suspicion of exposure may come from direct measures of indicator organisms, bathing water measures/status from a source external to the study (e.g. environment agency), or anecdotal or circumstantial information (e.g. proximity to likely point sources).

Justification: The group of recreational water users with the lowest pollution exposure in a study offers the most similar comparator between studies - even if in reality, the actual levels of exposure will differ at least slightly between studies. However, this is common in environmental health syntheses, and lowest vs highest/er comparisons combined with a

careful synthesis are regarded as a legitimate way of dealing with this issue (see PECO Example 2 in Morgan et al., 2018)*. Regarding the latter, our narrative and quantitative syntheses will include efforts to estimate potential heterogeneity arising from comparator/baseline differences, and account for this in drawing conclusions.*

Outcome

Estimated cases of infection: Confirmed or suspected cases of ear, eye, gastrointestinal, respiratory, skin, or urinary tract infection caused by microorganisms. These may be based either on self-reported cases of infection (such as gastrointestinal symptoms), diagnosis by a medical practitioner, and/or microbiological testing of human samples (the term 'estimated cases' captures the varying levels of certainty). When based on symptoms, reported outcomes in the paper will be considered eligible/classified into the six infection categories according to the schema presented in Table S2. When considering diagnostic and microbiological methods of estimating infections, it is important to note that both symptomatic and asymptomatic infections are eligible for inclusion – and relatedly, that our definition of 'infection' includes 'colonisation', as this is the first stage of the infection process (Dani, 2014).

Justification: In most studies, we anticipated that infection will rarely be established with absolute certainty, and instead estimated from self-reported symptoms. For the same reason, we do not specify that certain faecal microorganisms must be the agents of infection – especially as non-faecal microorganisms associated with wastewater and runoff may also play a role (especially in non-gastrointestinal infections). Nonetheless, some confirmatory evidence of infection to accompany symptomology is highly desirable, and we will account for this by 1) assigning lower risk of bias/higher certainty to such outcome measures in quality appraisal; 2) exploring the potential quantitative effect of the directness with which cases of infection are estimated in a dedicated sensitivity analysis.

Information sources

Recognising that librarian co-authors improve the quality and reporting of systematic review searches (Rethlefsen et al., 2015), our information sources (and search strategy, next section) have been selected in consultation with our librarian/information specialist co-author (AB). We selected from and expanded upon the information sources used in Leonard et al., (2018a), based on an analysis of whether they were useful information sources for finding the seawater studies included in this previous review (Supplementary Information 3; Figure S1). Broadly, they are grouped into bibliographic databases which will serve as our primary information sources, and grey literature sources which will serve as our secondary information sources.

Bibliographic databases

The bibliographic databases that we will search are MEDLINE (via Ovid), Web of Science (SCI, SSCI, AHCI, CPCI-S, CPCI-SSH, ESCI), Environment Complete (via EBSCOhost), and Global Health (via Ovid). Unlike in our previous review, EMBASE, BIOSIS and GreenFILE will not be searched, as a search summary table of the results indicated that they did not retrieve any unique results that were not found in other databases (Figure S1). The process for choosing the primary information sources is detailed in Supplementary Information 3. We will also use the citationchaser software (Haddaway et al., 2022) to perform forwards citation searching of the 2018 review (Leonard et al., 2018a) as well of the 40 included reports within it. All of the results obtained via these methods will be downloaded into EndNote (The EndNote Team, 2024).

Google Scholar

Google Scholar indexes scientific papers but cannot be searched reproducibly (Gusenbauer, 2019; Gusenbauer and Haddaway, 2020), and indexes research produced by both academic and non-academic institutions. Following PRISMA2020, we consider it both a bibliographic and grey literature 'database' (Rethlefsen and Page, 2022) that is worth searching despite its limitations, because it can capture results not captured in searches of select bibliographic databases (Haddaway et al., 2015).

Grey literature sources

We will also search grey literature sources, which we define according to the Luxembourg definition as "information produced on all levels of government, academia, business and industry in electronic and print formats not controlled by commercial publishing i.e., where publishing is not the primary activity of the producing body" (ICGL, 2014). In the context of this review, grey literature sources are specifically considered to be searchable websites of non-academic Global North organisations that have a vested interest in the relationship between environmental exposures and human health. The following fourteen organisational websites will be searched: 1) European Environment Agency (EU and non-EU member states and cooperating countries); 2) GOV.UK; 3) Environmental Protection Division (Georgia); 4) Ministry of Health (Ukraine); 5) Ministry of Environment (Moldova); 6) Health Canada (Canada); 7) Environmental Protection Agency (United States); 8) Japan Environment Agency (Japan); 9) Ministry of Environment (South Korea); 10) Water Quality Australia (Australia); 11) Ministry for the Environment/Manatū Mō Te Taiao (New Zealand); 12) Cabinet of Israel (Israel); 13) UN Environment Programme; 14) World Health Organisation. The process for choosing the grey literature sources is detailed in Supplementary Information 3. Due to resource constraints, it is only feasible to search these sources in English, but nonetheless these countries commonly produce English-language documents on water quality. As for the primary information sources, forwards and backwards citation searching will be performed on any new included papers that emerge from them.

Search strategy

Bibliographic databases

Our search strategy was also developed in consultation with our librarian/information specialist co-author (AB), who will also execute the strategy for the review. To develop this strategy, we tested and improved upon the search strategy used in Leonard et al., (2018a) to find studies with water/pollution exposures, before pilot testing the new strategy (Supplementary Information 4).

For searching the primary information sources (bibliographic databases and PQDT Global), the search terms used in the previous review have been amended based on the experience and results of Leonard et al., (2018a). This is to make our searches more efficient at finding all articles relevant to the review question, whilst minimising the number of irrelevant records retrieved from each database. The main changes were the addition of new search terms and the refinement of existing ones to avoid redundancy. The titles and abstracts of the seawater studies included in Leonard et al., (2018a) – and the 70 freshwater studies listed in the supplementary materials but not synthesised - have also guided the choice of new search terms, which now take into account any changes in language over the last 10 years. The process used for designing this search strategy is detailed in Supplementary Information 4, with an example search strategy for MEDLINE with these terms in Supplementary Information 5.

The new search strategy was pilot tested in MEDLINE on the basis of its ability to retrieve the 38 previously-included report of the 40 reports included in Leonard et al., (2018a) which were indexed in MEDLINE. The new search strategy retrieved all of the 38 previously included reports, compared to the 29 found using the previous search (Supplementary Information 4).

Given that the search strategy (and information sources) was developed based on Leonard et al., (2018a)'s seawater-focussed review, we also validated that the search strategy captured relevant freshwater studies. Among the 70 freshwater studies listed in the supplementary materials of Leonard et al., (2018a), we identified 9 studies with pollution exposures. All of these 9 studies were retrieved by our search.

Google Scholar

Given the large number of results a Google Scholar search typically produces, we will only search for seawater/freshwater/bathing water/recreational water terms in the titles of indexed reports, via Publish or Perish (Harzing, 2007). Pilot testing of this search strategy revealed a practicable number of hits including some of potential relevance to the review (Supplementary Information 4).

Grey literature sources

A different search strategy will be adopted for searching organisational websites, given the organisational websites included do not have search systems that are as sophisticated as those for the primary sources. Specifically, we intend to search for anything related to bathing or recreational water, filtering results on a more case-by-case basis depending on how manageable the number of records retrieved is. Example search strategies and numbers of hits for each organisational website detailed in Supplementary Information 5.

The new search strategy was pilot tested by first backcalculating the original sources of the 13 seawater reports included in Leonard et al., (2018a) but not identified in any of the bibliographic databases (Supplementary Information 4; Figure S1). Subsequently, we validated that our new search strategy captured the report in this suspected source in a similar manner how the primary information sources were validated (Supplementary Information 4; Figure S2).

Data management

Search outputs (titles, abstracts and other bibliographic information) will be downloaded from bibliographic databases as bibliography files (e.g. .ris, .bib files) and combined in EndNote (The EndNote Team, 2024). Duplicates will be removed via EndNote's deduplication tool before exporting as the de-duplicated bibliography file. This deduplicated file will be uploaded to the Rayyan systematic review management software (Ouzzani et al., 2016), in which we will attempt to further remove duplicates with a combination of automatic duplicate detection and manual checking, before it is screened.

Data from the studies will be stored in an Excel (Microsoft Corporation, 2018) spreadsheet/comma separated values files on a shared university OneDrive during the review, with data and code necessary to reproduce the meta-analyses (though not raw data from others' studies) published alongside the final review paper.

Selection process

Titles and abstracts (and subsequently selected full texts) of the deduplicated reports will be double-blind screened in Rayyan using the PECO-structured eligibility criteria. MLJ will screen all reports (acting as primary reviewer) and one of four second reviewers will screen a random but unique subset of all the reports (created using the 'Create Sample' feature of Rayyan) blind to MLJ's decisions. For example, if there were 1000 reports to screen, MLJ would screen all 1000 whereas each of the four reviewers would screen a unique subset of 250 of these, in order to ensure that all reports are double-screened.

Before beginning screening, the ability of all five reviewers to consistently apply the eligibility criteria will be validated by pilot testing in which each of the five reviewers screened a random set of 100 reports. Based on this, we will identify opportunities to make minor clarifications to the review criteria in order to minimise conflicts between reviewers at the end of each screening stage (these will be reported in the final review). Conflicting decisions on the

inclusion or exclusion of records will be resolved by discussion to try to reach a consensus, with a third reviewer brought in to resolve remaining discrepancies.

Data collection process

Data will be extracted by MLJ to an Excel (Microsoft Corporation, 2018) spreadsheet/comma separated values file in the first instance, stored on the university OneDrive which will also produce a version-controlled backup of the data. PDFs will be highlighted to indicate where extracted data has come from.

Data will be checked by a second reviewer, although this will not be blinded for practical reasons. Primarily we will extract data from published reports. Where a contact is available, we will contact the author(s) of the included studies for study data in order to obtain more precise estimates where needed. Authors will be emailed using the corresponding author's address and/or other authors' addresses and asked to share their data for the purposes of inclusion in a systematic review/meta-analysis, with at least one follow-up email if there is no response. A second reviewer will check the data extraction and any discrepancies resolved through discussion, and if necessary, after consultation with another member of the review team.

Data items

We intend to extract the following data items for each study using information available in reports/from authors, though the exact list is subject to evolve during the data extraction process (e.g. there may be sub-categories to each of these items, additional items extracted):

Study name (published name or arbitrary unique identifier); study funder; year(s) of study period; country of study; study design; population studied; selection criteria for study population; type of open water (seawater/freshwater/transitional); sub-type of open water (e.g. coastal, estuary, river, lake); type of recreational use of water studied (e.g. swimming, surfing); description of each exposure and comparator group; method of assignment to each exposure and comparator group; number of participants originally assigned to each exposure and comparator; number of participants analysed in each exposure and comparator group (and reasons for loss to follow up); actual or suspected source of pollution in each exposure and comparator group; method of characterising the level of pollution in each exposure and comparator group (e.g. indicator microorganism, water quality status, rainfall level) in each exposure and comparator group; level of pollution in each exposure and comparator group; outcome assessment method (e.g. self-reported, diagnosis by a medical practitioner, microbiological testing); type of infection assessed in the outcome (e.g. any, ear, gastrointestinal); number of cases in each exposure and comparator group; number of non-cases in each exposure and comparator group; estimate of association/effect size type (e.g. risk ratio, odds ratio); crude value and confidence intervals of estimate of association/effect size type; adjusted value and confidence intervals of estimate of association/effect size type; covariates adjusted for; statistical methods used (e.g. model type); bibliographic information (e.g. authors, year of publication, DOI) of report in which outcome is reported.

Outcomes and prioritisation

The primary outcomes for which data will be sought are data comparing estimated cases of different types of infections in open water users exposed to the lowest level of pollution in a study versus those exposed to higher levels of pollution in a study (i.e. pollution exposures) – typically expressed as the number of cases/non-cases in each study arm or an estimate of association/effect size (e.g. risk ratio, odds ratio). The six types of infection for which data will be sought are infections of the ear, eye, gastrointestinal tract, respiratory tract, skin, and urinary tract. Ideally, published reports will report a maximum of one outcome related to each of our six categories, with a clear indication of the category to which each reported outcome belongs (ideally based on diagnostic or microbiological confirmation). In practice however, this is not always the case, and typically outcomes are reported based on categories of (self-reported) symptoms, with multiple outcomes from each category reported, and/or it is not always clear to which of our six categories each reported outcome belongs. To deal with this, we will use a pre-defined categorisation system based on NHS symptomatology information to categorise symptom outcomes into each of our six types of infection (Table S2). If multiple outcomes per infection type exist, these will be aggregated into a single estimate of association per infection type for the narrative synthesis, or accounted for through the random effects structure of our meta-analytical models for the quantitative synthesis.

Our previous review (Leonard et al., 2018a) used a more elaborate classification system for infection types/symptoms. This system included different sub-categories of symptoms within different types of infection and different case definitions within those, resulting in 10 categories emerging from symptoms of any type, the ear, and the gastrointestinal tract. We consider it undesirable to use this classification system in this review, for several reasons. Firstly, dividing the meta-analytical dataset between many models/levels of a moderator in a model further depletes the already limited statistical power typical of this type of meta-analysis. Secondly, our previous pooled estimates of the association/effect for different case definitions of the same symptom category/similar types of symptoms were similar - suggesting that at least some of the original sub-divisions were unnecessary. Finally, the previous classification system complicates the interpretation the results for those wanting to use them in an applied context (e.g. it is not immediately clear which pooled estimates of the association/effect for each type of infection is most relevant). This third problem would be exacerbated by the addition of 4 new sites of infections in this review, and the accompanying factorial increase in the number of sub-categories. For all these reasons, we choose to use a simpler, infection type-based classification system in this review, which has only 6 categories.

Bias in individual studies

We will assess risk of bias in the included studies using Cochrane's ROBINS-E and RoB 2 tools for observational and randomised studies, respectively (Higgins et al., 2024; Sterne et al., 2019). Risk of bias will be assessed at the exposure-outcome level, meaning that if a study

contains multiple relevant exposures and outcomes (e.g. infection types), each relevant outcome within each relevant exposure will be assessed. Two reviewers will independently assess risk of bias for each study following the guidance notes available for each tool. Any discrepancies between the two reviewers in risk of bias judgements for each domain will be resolved through mediation by a third reviewer. If a consensus cannot be reached, we will conservatively defer to the highest risk of bias judgement made by the three reviewers.

ROBINS-E

Most studies in this field are non-randomised, due to the difficulties of running randomised trials to assess the effects of environmental exposures such as exposure to (polluted) open water. Risk of bias of this majority of non-randomised studies will be assessed using the ROBINS-E tool (Higgins et al., 2024), implemented using the Microsoft Excel template available at riskofbias.info. These studies will be judged as being of 'Low', 'Moderate', 'Serious' or 'Critical' risk of bias (or 'No information') in each domain and overall, in line with the tool's schema.

RoB 2

There are at least three randomised studies that include pollution exposures (Kay et al., 1994; Wiedenmann et al., 2006; Fleisher et al., 2010). Risk of bias of this minority of randomised studies will be assessed using the RoB 2 tool (Higgins et al., 2024; Sterne et al., 2019), implemented using the interactive Microsoft Excel worksheet available at riskofbias.info. These studies will be judged as being at 'Low' risk of bias, having 'Some concerns', or as being at 'High' risk of bias in each domain and overall, in line with the tool's schema.

When assessing deviations from intended interventions (RoB 2 Domain 2), studies will be assessed for the 'per protocol' effect of adhering to the polluted seawater exposure intervention, rather than the 'intention-to-treat' effect. This is because:

1. These trials are not meant to be representative of a real-life medical intervention. Controlled exposure to (polluted) seawater in these trails is being used as a proxy to study the effects of real-life *exposure* to seawater, which typically takes place in a an entirely self-controlled recreational context, rather than a clinician-mediated medical one. Therefore, the effects of 'non-adherence to therapy' are less relevant.
2. Related to the above, individuals and other stakeholders in this field are most often interested in the effect on those who *actually* expose themselves to open water, because it most closely relates to the implications of their choice to expose themselves/advise against exposure to polluted seawater. In this context, 'per-protocol' analysis may more directly inform individual and policy-level decision making (Higgins et al., 2019 p5-6; J. P. Higgins et al., 2024).
3. Finally, and perhaps most importantly, the hypothesised effect of exposure to polluted seawater is a harm/adverse one, and so the danger of underestimating the

> effect arguably outweighs the danger of overestimating it. Therefore 'per-protocol' analysis is arguably the more conservative choice here - rather than 'intention-to-treat' analysis, which is expected to underestimate the intervention effect in a blinded, placebo-controlled trial where there is non-adherence to assigned interventions (Hernán and Hernández-Díaz, 2012; Higgins et al., 2019 p6).

Given naïve 'per-protocol' and 'as-treated' analyses can be problematic because they do not adjust for prognostic factors that may influence whether individuals receive or adhere to their assigned intervention, we will assess the 'per-protocol' effect as rigorously as possible. Specifically, using the information presented in trial reports, we will attempt to assess the occurrence of non-protocol interventions (Additional signalling question 2.3), failures in implementing the intervention that could have affected the outcome (Additional signalling question 2.4), and non-adherence to the assigned intervention by trial participants (Additional signalling question 2.5). If evidence of deviation/non-adherence from the intended intervention is found, we will then assess whether study authors used an appropriate analysis to assess adherence to the intervention (Additional signalling question 2.6)

Piloting of the two tools

Use of ROBINS-E and RoB 2 for assessing these studies was piloted by two reviewers (MLJ and EL) by assessing one exposure-outcome combination from one non-randomised and 1 randomised study from the 9 included studies (20%) with a pollution exposure assessment in the original review (7 listed in the Study Characteristics table and a further 2 included in the seawater exposure synthesis but which also included a polluted seawater exposure assessment – see Supplementary Information 2).

To begin the piloting process, an initial meeting was held between MLJ and EL to review the use of both tools in the context of this body of literature, and clarify any initial uncertainties regarding the application of the tools. Subsequently MLJ and EL independently assessed one exposure-outcome combination (estimated cases of gastrointestinal infection) for each of the 2 reports, and then met to discuss and resolve any discrepancies in risk of bias assessments for each domain. All discrepancies were resolvable and so the introduction of a third reviewer was not necessary for these 2 reports. Nonetheless, the process provided a helpful opportunity for training of both reviewers (MLJ had not previously used ROBINS-E and EL had not previously conducted risk of bias assessments) and trial their implementation for studies in this field (Supplementary Information 6).

Risk of bias results will be visualised as 'traffic light' plots, which we consider to offer the best study-level visualisation of the results. These plots will be produced using the 'robvis' R package (McGuinness and Higgins, 2021). Risk of bias judgements will be used in a sensitivity analysis testing whether risk of bias is driving the results in the primary analyses, as well as in certainty assessment via GRADE.

Synthesis

Hypotheses and estimate of the association/effect

The hypotheses we will explore in our syntheses is:

H1. In recreational open water users, there will be a positive association between confirmed or suspected exposure to water polluted by wastewater and run-off, and estimated cases of infection by microorganisms

This hypothesis is formulated as an explicit prediction based on our overarching PECO-structured research question (Table 1). This hypothesis will be tested 6 times - once for each of the 6 infection types - for each of the narrative and quantitative syntheses (though only the latter is a formal test, and hence will be corrected for multiple testing – see below). Risk ratios representing the difference in estimated cases of infection between the participant group of recreational water users with the lowest exposure to wastewater (including sewage) and run-off pollution in a study, and the group(s) with higher exposure in a study will be used to assess evidence for/against our hypotheses.

Risk ratios were selected as the estimate of association because of their collapsibility (meaning that their size does not change if adjustment is made for a variable that is not a confounder; Cummings, 2009) and associated with this, their greater interpretability (relative to odds ratios which are more prone to misinterpretation associated with overestimating risk; Cummings, 2009; Grant, 2014a). The potentially includable studies identified during scoping (Supplementary Information 2) also frequently report associations/effects as risk ratios and/or cases/non-cases in the exposure and comparator groups. Adjusted risk ratios are the preferred metric in both our qualitative and quantitative syntheses, but we will seek to extract and make use of both where possible (see sensitivity analyses). This is because whilst there is a strong argument for adjusting effect estimates for confounding (particularly in non-randomised studies which we anticipated will make up the majority of studies in our synthesis), there is also a reasonable counter-argument that adjusted effect sizes are less comparable because each study adjusts for different confounders (Higgins et al., 2023; Reeves et al., 2023). We will attempt to derive crude and/or adjusted risk ratios either by: 1) directly extracting crude and/or adjusted risk ratios and 95% confidence intervals from reports; 2) extracting alternative estimates of association (e.g. odds ratios) from reports and converting them to risk ratios using the average baseline risk in the comparator/control group (Grant, 2014a, 2014b; Higgins et al., 2023); 3) extracting or obtaining from authors the estimated numbers of cases of an incident infection in the exposure groups and comparator/control group considered from reports, and then using them to calculate a crude risk ratios and their 95% confidence interval. Sometimes risk ratios/odds ratios comparing the least exposed group to more exposed groups will not be presented in the reports themselves (for example because groups with different levels of exposure to pollution are instead compared to the non-bather group e.g. Kay et al., 1994). In such cases, we will attempt to (re)calculate risk ratios comparing the least with the more exposed groups, using the case number information extracted from reports/obtained from authors.

A key decision is to use risk ratios comparing recreational water users with the 'lowest' suspected exposure to pollution in a study, to recreational water users with 'higher' suspected exposure in a study. We have already outlined our justification for focussing on estimates comparing recreational water users with one another (i.e. of the effect of pollution exposure), rather than with non-recreational water users (i.e. of the effect of water exposures). Nonetheless, pollution exposures might seem a more subjective to some than water exposures, given that:

1. The cut-off for lowest/higher pollution exposure is less defined than for non-bather/bather, and hence probably varies more between studies (heterogeneity in the exposure). Associated with this, there may be multiple 'higher' groups (though this is intended to be dealt with through grouping in the narrative synthesis and explicitly explored through random effects and dose-response meta-regressions in the quantitative synthesis).
2. Exposure to polluted water cannot typically be established with the same degree of certainty as exposure to water *per se* (error in the measurement of the exposure)
3. Whereas in non-bather/bather comparisons there is only one opportunity for heterogeneity and/or measurement error to affect the estimate, in lowest /higher pollution bather comparisons there are two (i.e. the comparator is affected by heterogeneity and error in the measurement of the pollution in the comparator, as well as the exposure)

We consider these important limitations of focussing on pollution exposures. Nonetheless, to some extent these are characteristics of the studies being synthesised, rather than the synthesis *per se*. For syntheses of environmental exposures like wastewater and run-off pollution - which resist straightforward characterisation or for which the choice of cut-off value is unclear - there is arguably little choice but to define exposures/comparators in the relatively broad way it is defined in the literature (see PECO Scenario 2 in Morgan et al., 2018). Of course, this is likely to result in greater heterogeneity in the estimates included in the synthesis, which might be seen as particularly problematic for the meta-analysis stage of the review. However, given the state of the literature, we believe that trying to use eligibility criteria to eliminate the subjectivity with which these studies define their exposure and comparator groups is undesirable. Doing so risks excluding studies based on relatively arbitrary cut-offs that may not turn out to be consequential for the outcome, and does not allow the review sufficient scope to establish whether sensible cut-offs that are consequential for the outcome do exist. Hence, we believe it is better to complement looser eligibility criteria with a more diligent and in-depth execution of all stages of a systematic review (not just meta-analysis). This allows narrative synthesis, risk of bias assessments, subgroup and sensitivity analyses (targeted at addressing subjectivity in the comparison) to contextualise the meta-analytical results - brought together in a certainty of evidence (e.g. GRADE assessment) assessment that allows readers to carefully interpret the quantitative results. Thus, whilst our review is likely to be focussed on a narrower set of studies with a less consistent definition of exposure than previous reviews comparing non-recreational water users and recreational water users, this will be offset by its focus on a comparison more

relevant to addressing the role of pollution and a more detailed assessment of the evidence and its uncertainty.

Qualitative synthesis (narrative synthesis)

A narrative synthesis (supported by a study characteristics table) will be used to describe the characteristics and relevant reported results of all included studies. The intended methods for this synthesis are described below, following the relevant (protocol-stage) parts of the SWiM reporting guidelines (Campbell et al., 2020; Table S3).

Separate narrative (and quantitative) synthesis will be conducted for seawater, freshwater, and transitional water (e.g. estuaries, fjords), though some comparison of the results will be made in the Discussion. Within this, we will group studies in a way which facilitates easier digestion of the narrative synthesis. 'Grouping' refers to the way in which we will group exposure-outcome combinations when narratively discussing them in the study characteristics table and main text (Campbell et al., 2020). The main grouping factor in our narrative synthesis (reflected in sub-tables in the study characteristics table and sub-headings in the main text) will be infection type, given that we expect pollution to have a stronger effect on some infection types than others. Within this main outcome-level grouping factor, outcomes will be further grouped and discussed by method of estimating cases of infection (e.g. self-reported symptoms, diagnosis by a medical practitioner, microbiological testing) - given that diagnosis and microbiological confirmation are expected to produce more accurate estimates of cases of infection. Exposures will be grouped by the directness of method of measuring/inferring pollution (e.g. anecdotal/circumstantial information, water quality information, direct measures of indirect organisms) and level of pollution, given that the method of characterising the exposure and the level of pollution are hypothesised to be key influences on the estimated effect, respectively. tWe are not primarily interested in differences between populations, so we do not anticipate grouping our synthesis by population - though if there are multiple populations within a study, we will present the exposure-outcome level estimate derived from meta-analytical aggregation (Supplementary Information 9). At the top-level, studies will be grouped by whether they are non-randomised or randomised, in line with our prioritisation of the latter for drawing conclusions (see below).

We will present will be the crude and adjusted risk ratios (and their 95% confidence intervals) as the standardised outcome metrics for each exposure-outcome combination, with adjusted risk ratios prioritised for drawing conclusions. For adjusted risk ratios, we will also present the confounders that have been adjusted for alongside the adjusted risk ratios in the study characteristics table. If studies provide multiple estimates per exposure-outcome combination (e.g. due to presentation of the results at the level of subgroups), we will present the exposure-outcome level estimate derived from meta-analytical aggregation (Supplementary Information 9). If risk ratios cannot be derived for any of the exposure - outcome combinations, we will briefly narratively describe the result for each exposure-outcome combination in the study characteristics table, based on the information presented in the report.

Narrative synthesis in the main text of the review will consist of summarising effect estimates (range and distribution of observed effects) as well as more explicit vote counting based on the direction of the effects (not the size or significance of results; McKenzie et al., 2023). Whilst vote counting is a limited synthesis method (Grainger et al., 2022), we contend that all narrative synthesis resorts to some form of vote counting, and there is very little alternative for summarising all included studies when some studies are not amenable to meta-analysis. At the very least, vote counting provides a basis for discussing the alignment or misalignment between reported and meta-analytical results; a benchmark for comparing if the results of studies well-reported enough for quantitative synthesis align with those that are not (which could be a form of publication bias). Due to the substantial risk of confounding and nocebo effects in this field (see Fleisher and Kay, 2006), randomised controlled trials will be prioritised when drawing conclusions from this synthesis by means of highlighting their results in contrast those of non-randomised studies. For similar reasons, estimates of the association/effect judged to be at the lowest risk of bias will also be prioritised through highlighting them in comparison to those judged to be at higher risk of bias. Narrative exploration of heterogeneity in effects will be kept to a minimum as advised (Campbell et al., 2020), and will consist only of ordering the study characteristics table by study design (complementing the narrative highlighting/contrasting of results from randomised and non-randomised studies) and exposure-outcome combinations within studies in ascending order of the level of pollution (which is hypothesised to be a key moderator of the effect). Certainty of the evidence included in the narrative synthesis will be assessed alongside that included in the quantitative synthesis, as detailed in the 'Confidence in cumulative evidence' section of this protocol.

As described above, study characteristics and results will be presented in a study characteristics table grouped following the grouping system. This table will present each study's design, populations, exposures, controls/comparators, outcomes, and the risk of bias judgement for each exposure-outcome combination (as well as the bibliographic references for the reports in which each exposure-outcome combination is reported). Additionally, we also intend describe any potential links these studies make between these associations/effects and 1) the effects of climate change; 2) the acquisition of antimicrobial resistant microorganisms by study subjects. This is on the basis of concerns about interactions between infections by microorganisms and climate change (Gogien et al., 2023; Hensyl et al., 2024; OFWAT, 2008) and antimicrobial resistance (Honda et al., 2020; Tipper et al., 2024), which are a focal point for the wider project of which this review is part (Horizon Europe Grant Agreement No 101057764).

Quantitative synthesis (meta-analyses)

Data synthesis criteria

As well including them in the qualitative synthesis (narrative synthesis), we will seek to include each study included in the wider review in a quantitative synthesis (meta-analyses). A study will be included in the quantitative synthesis provided at least one risk ratio can be obtained

from it following the process described for the narrative synthesis. As in our narrative synthesis, multiple exposure-outcome combinations are eligible for inclusion in our quantitative synthesis. Furthermore, multiple risk ratios per exposure-outcome combination are eligible for inclusion (e.g. repeated measures for the same exposure-outcome combination; risk ratios reported separately for different demographic subgroups of the study population). These relatively relaxed data synthesis criteria prevent the need to arbitrarily select a single risk ratio from several meeting the PECO criteria in each study, as well as reduce the loss of information associated with aggregating multiple risk ratios into one (which is especially useful to conserve for additional and alternative analyses e.g. of dose-response effects). The resultingly unbalanced number of risk ratios per study included in our meta-analytical models will be handled through the random effects included in our mixed-effects models (see below). Finally, we will include self-reported, clinician-diagnosed and microbiologically deduced infections in the same models, in order to maximise their statistical power. This may introduce heterogeneity, but this will be explored using a dedicated subgroup analysis based on method of outcome assessment.

Summary measures

General approach (response variables and random effects)

Seawater, freshwater, and transitional water data will be modelled (i.e. synthesised) separately for all primary, additional and alternative analyses. All models included in our primary, additional and alternative analyses will be built using the 'rma.mv' function of the metafor R package (Viechtbauer, 2024, 2010).

The random effects included in this model (specified via the 'random' argument in 'rma.mv'):

1. **Study ID** to account for the study-level effect.
2. **Time of sampling nested within exposure** to account for temporal autocorrelation between repeated measures of outcomes within the same exposure. This will be modelled with a continuous-time autoregressive structure ("struct = 'CAR'" option in rma.mv). More information about this approach can be found in Viechtbauer (2024), with Stasielowicz (2022) offering an example of its implementation.
3. **Unit-level observation ID** to account for residual within-study variance. This is to account for any residual variance between multiple effect estimates from the same study, after accounting for the effects of study (e.g. due to different of infection types/sub-types, or case definitions within the same study). Examples of this approach and its implementation can be found in Konstantopoulos (2011), Credé, Roch and Kieszczynka (2010), Viechtbauerr (2022), Viechtbauer (2007a).

The response variables included in these models will be log risk ratios and their variances (specified via the 'yi' and 'vi' argument in 'rma.mv'), transformed from the risk ratios and their 95% confidence interval derived the reports. Adjusted (log) risk ratios and their variances will be preferentially chosen over the crude (log) risk ratio where both are available, since these

adjust for confounders (with the exception of a sensitivity analysis explicitly exploring the consequences of this modelling choice; see Alternative analyses: Alternative modelling approaches). Variance will be modelled as a variance-covariance matrix accounting for shared-control non-independence, following Lajeunesse (2011) and Moran et al. (2021). This is to account for non-independence between multiple estimates of the association/effect based on the same comparator group, which emerge from studies with multiple exposure arms but only a single comparator arm (Lajeunesse, 2011; Noble et al., 2017; Nakagawa et al., 2023).

Primary analysis

As an initial test of our hypothesis, we will build a random effects model (Model H1a) including the random effects already described. This 'intercept-only' model will assess the direction, strength and variability of the overall effect. By characterising overall effect before accounting for infection type in this way, this model acts as a precursor to testing the hypothesis formally for each infection type.

For formally testing our hypothesis, we will build a mixed-effects models (i.e. meta-regression or a multivariate/multilevel model; Model H1b) including the random effects already described, as well as a fixed effect representing the six types of infection. This differs from the approach used in Leonard et al., (2018a) in that it uses one (mixed-effects) model for all infection types/symptom categories, instead of one (random effects) model for each. We adopt this approach here because it statistically preferable under most scenarios (Rubio-Aparicio et al., 2020, 2017). Furthermore, it simplifies the implementation and communication of the models and their results, since there are fewer models to build and present - especially when one considers that the number of sensitivity and subgroup factorially multiply from the number of primary models.

Interpretative criteria

The pooled estimate for each infection type (i.e. the fixed effect/predictor/moderator estimates from the meta-regression) from these models will be back-transformed from a log risk ratio to an easier-to-interpret risk ratio, before being used to assess the direction, strength and variability of the effects. When interpreting these effects, the first indicator for the presence of an effect that we will use will be the p-value (significance threshold of 0.05). However, in order to contextualise this result, we will also consider the degree to which the point estimate and 95% confidence interval of each pooled estimate surpasses negative effect (less than 1)/no effect (1)/positive (more than 1) risk ratio thresholds (i.e. strength of the association/effect) – especially where p-values are only marginally insignificant.

Residual heterogeneity will be characterised using the Q_e test for heterogeneity, pseudo-I^2 statistic sensu Jackson, White and Riley (2012), and the 95% prediction intervals of each estimate. The Q_e test for heterogeneity and pseudo-I^2 will be used to assess residual heterogeneity at the level of the whole model (i.e. to what extent random effects - and in the

case of Model H1b, fixed effects - explain differences between individual estimates). The prediction interval will be used to assess heterogeneity at the level of the outcome (i.e. all outcomes for Model H1a and each type of infection for Model H1b). This is following the recommendation of IntHout et al., (2016), who argues that prediction intervals are an easier-to-interpret metric of heterogeneity than τ^2 or I^2 (which are also unreliable in small meta-analyses; von Hippel, 2015). Prediction intervals are also a far more convenient-to-generate metric of heterogeneity for individual levels of a moderator.

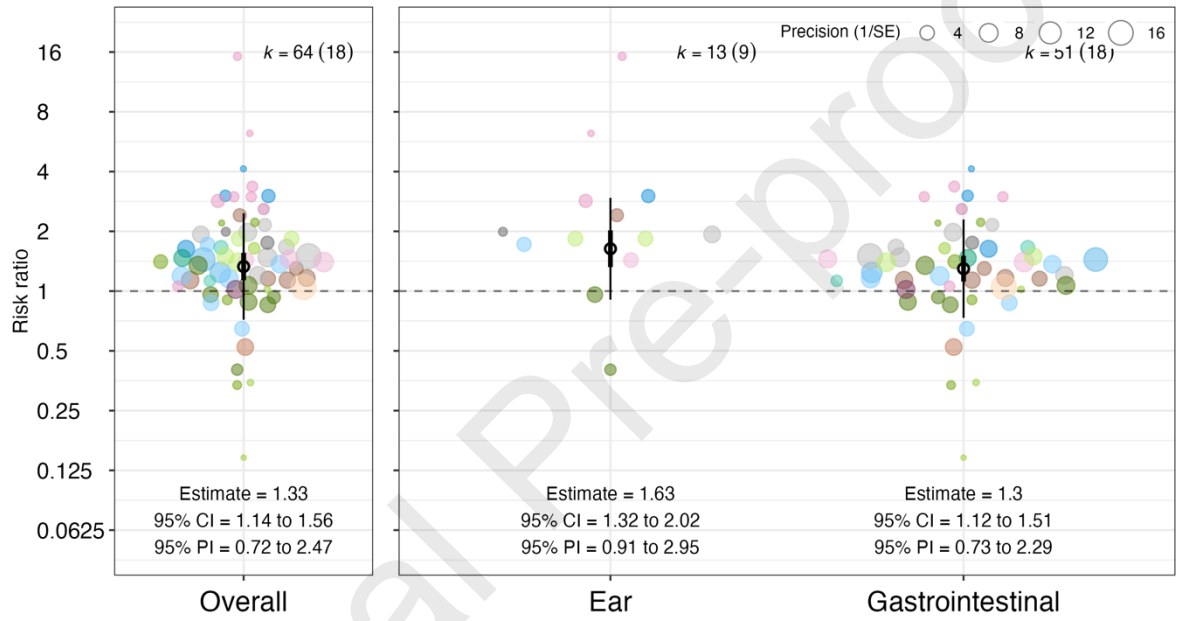
For Model H1b, to understand whether type of infection is generally is good at explaining variance in the effect across studies (i.e. whether the strength of the association/effect differs between infection types), we will use the Q_m test of moderators and the pseudo-R^2 statistic of the model (Viechtbauer, 2022).

Model fit overall will be assessed using profile likelihood plots of the variance components of the models, as well as sensitivity analyses (see 'Additional analyses').

Visualisation

The two models will be visualised orchard plots (Nakagawa et al., 2021) showing the overall and per-infection type pooled risk ratio estimates. These visualisations helpfully and succinctly display not only the point estimate and confidence interval, but also the prediction intervals and individual risk ratios underpinning each pooled estimate. The plot will display this information on a log risk ratio-scale but with annotations on a risk ratio scale to facilitate interpretation (Figure 1).

Figure 1: *Example orchard plots demonstrating the meta-analytical (left) and mixed-effects models approach using the (sea)water exposure data from members of this groups' previous review (Leonard et al., 2018a, 2018b). This plot is based on the results of a random-effects model (i.e. meta-regression or a multivariate/multilevel model) using infection type the moderator (right). In these plots, the pooled overall or per-infection type risk ratio is represented by the 'trunk' of the tree (black circle), with its 95% confidence interval represented its 'branch' (shorter, thicker line through it), and its prediction interval represented by its 'twig' (longer, thinner line through it). Individual points represent the risk ratios for each exposure -outcome combination from which these pooled estimates are derived, coloured according to the study from which they are derived. Risk ratio information is displayed on a log scale.*



Additional analyses

Sensitivity analyses

Risk of bias

Estimates of the association/effect deemed to be at high risk of bias may be less reliable and lead to misleading conclusions if these estimates are driving meta-analytical results. Accordingly, The Cochrane Handbook of Systematic Reviews of Interventions recommends restricting primary analyses to estimates judged to be at the low risk of bias where possible (Boutron et al., 2024). However, for systematic reviews of environmental exposures this is rarely feasible, given the number of included studies is typically small relative to that in systematic reviews of clinical interventions (and the number of low risk of bias studies is even smaller). In such circumstances, restricting the primary analysis to the lowest risk of bias studies or stratifying it into multiple analyses (which also has other problems; Boutron et al., 2024) is likely to result in a primary analysis that is underpowered (Cuijpers et al., 2021). Furthermore, it may be unnecessarily underpowered, since even when conducted rigorously, risk of judgements are to some extent necessarily subjective (Frampton et al., 2022).

Therefore, instead of restricting out primary analysis to low risk of bias studies, our first, most important sensitivity analysis will explicitly test whether high risk of bias studies are driving the results of our primary analysis, and adjust estimates accordingly. Specifically, we will remove studies with a high risk of bias (High risk of bias in RoB 2 or High risk of bias/Very high risk of bias in ROBINS-E), before re-running Model H1b (see Hill et al., 2022 and Izcovich et al., 2022 for similar sensitivity analyses). For each infection type, we will then compare the risk of bias-uncorrected and risk of bias-corrected pooled estimates to understand the extent to which high risk of bias studies drive the results in our primary analysis. For infection types for which the confidence intervals of the uncorrected and corrected estimates do not overlap, we will remove high risk of bias studies for all downstream analyses. The more consequential nature of this sensitivity analysis for the systematic review reflects the importance of accounting for risk of bias emphasised in the Cochrane Handbook (Boutron et al., 2024), as well as previous cases of high risk of bias studies apparently driving misleading conclusions in systematic reviews (Hill et al., 2022; Izcovich et al., 2022). An additional, more conservative check on risk of bias is also built into our GRADE assessment in the certainty of the estimates, providing an additional opportunity to account for the (potential) effects of risk of bias when drawing conclusions.

Directness of the indicator of pollution exposure

Another factor that could have an important bearing on the conclusions of our quantitative analysis is the directness of the indicator used to establish pollution exposure in participant groups. Theoretically, the lack of an effect in the least confirmed exposures (where participant exposure to wastewater and runoff may not actually have taken place) could obscure an

effect in the most confirmed exposures to wastewater and runoff, causing us to mistakenly accept the null hypothesis of no effect. This should partly be addressed in risk of bias assessments' domains covering characterisation of the exposure, but nonetheless given its potential importance we will also conduct a dedicated sensitivity analysis to try to quantitatively validate that such a phenomenon is not biasing our results. This will consist of restricting our analysis to exposure-outcome combinations where the investigators have based the exposure category on environmental microbiology directly collected from the site of purported exposure. This will usually be data collected by the study investigators themselves as part of the study, but could also come from environmental agencies if the data is of a high enough spatial and temporal resolution to suggest exposure of all participants in the 'exposed' group.

Directness of the indicator of infection

Similarly to the indicator of exposure, the indirectness of the indicator of infection could affect the reliability of our quantitative conclusions. It is our impression that most studies in this literature rely on self-reported symptoms, which could either result in under- or over-estimation of the number of cases of infection (Fleisher and Kay, 2006). In this sensitivity analysis we will therefore restrict our data to exposure-outcome combinations that have been confirmed by a clinician or microbiological testing. This will provide some indicator of whether pooled estimates of association are different for confirmed cases of infection.

Leave-one-out-analysis

We will conduct a standard 'leave-one-out' analysis, in which one study is removed at a time, rerunning Model H1b and comparing 'before' and 'after' estimates each time. This sensitivity analysis attempts to check for the presence of influential studies, recognising that certain studies may have a disproportionate effect on the pooled effect(s).

Replacing adjusted risk ratios with crude risk ratios

The third planned sensitivity analysis is an analysis that seeks to test the extent to which adjusted risk ratios are driving the pooled estimates. The preferential use of adjusted risk ratios in our primary analysis is motivated by the fact that adjusted risk ratios attempt to account for confounders. However, a critique of this approach is that this is not appropriate to combine these risk ratios because studies typically adjust for different sets of covariates/confounders, reducing their comparability. Therefore, in this sensitivity analysis we intend to replace adjusted risk ratios with crude risk ratios where available/calculable, before rerunning model H1b and comparing the estimates.

Assuming different degrees of correlation between effect sizes from the same study

The fourth planned sensitivity analysis is an analysis assuming different degrees of correlation between risk ratios from the same study. Aside from repeated measures (accounted for via a random effect), studies may contribute multiple risk ratios to our meta-analyses if they report estimates of the association/effect at the level of subgroups (e.g. demographic subgroups, sub-outcomes). These risk ratios are not statistically independent since they come from the same study. Whilst this non-independence will be somewhat accounted for already by our model's random effect estimates for observations nested within each study, one can also explicitly specify different degrees of assumed correlation between estimates of association from the same study by fitting the model as a variance-covariance matrix assuming different degrees of independence (sensu O'Dea *et al.*, 2019). Following this approach, we will assume correlations of 0.2, 0.4, 0.6, 0.8 in order to see how sensitive the detected effects in our primary analyses are to differing degrees of assumed correlation.

Subgroup analyses

Subgroup analyses will be used to explore factors that potentially explain the residual heterogeneity observed in the pooled estimates for each infection type, after removing high risk of bias estimates where necessary. Subgroup analyses will consist of splitting the dataset into subgroups before re-running Model H1b for each subgroup. Given the limited statistical power of subgroup analyses, we will follow advice to limit subgroup analysis to a small number of pre-specified analyses with a limited number of subgroups within them (Burke et al., 2015; Cuijpers et al., 2021). These analyses will focus on trying to deduce the extent to which potential issues in the methods used to

Specifically, in three subgroup analyses we will split and model the dataset according to:

1. **Suspected of being exposed to human waste pollution (Exposed to human waste/Not exposed to human waste)**. Human waste (e.g. faeces) is likely to be infectious than animal waste. This subgroup analysis will therefore compare meta-analytical results according to whether the authors report that the pollution exposure was likely to include human waste. If this is unclear, we will omit the data from the analysis.
2. **Predominant source of pollution (Run-off/untreated wastewater/treated wastewater).** The type of pollution may affect the estimate of the association, due to differences in volume, frequency and content (Gasperi et al., 2010; Pistocchi, 2020; Environment Agency, 2023b). Perhaps the most basic hypothesis is that the largest effect will be seen in untreated wastewater, as is commonly assumed in public and policy discussions of this issue. This subgroup analysis will therefore investigate the potential effect of the predominant source of pollution in the exposure reported by investigators. If this is unclear, we will omit the data from the analysis.
3. **Type of infectious agent (virus/bacteria/fungi/eukaryote)**. Where there is microbiological confirmation of the agent of infection, it is potentially informative to

split estimates of the association according to broad taxonomic division of the infectious agent. This could help understand which broad group is likely to be causing the most infections for each infection type, for example – especially when paired with dose-response investigation of levels of these microorganisms in the water (see below). This subgroup analysis will therefore split the meta-analytical dataset according to broad taxonomic unit.

These analyses will primarily be exploratory, again in recognition of the limited statistical power of subgroup analyses in meta-analyses (Cuijpers et al., 2021). However, if there is evidence that residual heterogeneity in Model 1b is substantially reduced within certain subgroups, then these analyses will also inform the choice of estimates for summarising the findings of the review (see 'Confidence in cumulative evidence'). Specifically, we will consider whether one or more of the subgroups containing at least two studies (the minimum needed for a 'meta-analysis') moves the original pseudo-I^2 estimate from a higher to a lower Cochrane interpretative category (Deeks et al., 2024). If this is the case, then we will assess certainty for these subgroup estimates and present in them in the Summary of Findings table (instead of assessing the original pooled estimates). For example:

- If the pseudo-I^2 for Model 1b in the primary analysis was 20%, then certainty would be assessed for the original pooled estimates for each type of infection. This is because this falls within the "0% to 40%: might not be important" interpretive category of Deeks et al., (2024), implying there is limited evidence for residual heterogeneity in the first place.
- If the pseudo-I^2 for Model 1b in the primary analysis was 80% ("50% to 90%: may represent substantial heterogeneity"), but none of the subgroup models had a pseudo-I^2 below 50%, then certainty would be assessed for the original pooled estimates for each type of infection. This is because although the pseudo-I^2 is reduced, it stays within the "50% to 90%: may represent substantial heterogeneity" interpretative category of Deeks et al., (2024).
- If the pseudo-I^2 for Model 1b in the primary analysis was 80% ("50% to 90%: may represent substantial heterogeneity"), but the pseudo-I^2 for the subgroup model for 'Exposed to human waste' (from subgroup analysis 1) was 40%, then for each type of infection, certainty would be assessed for estimates related to "Exposed to human waste". This is because the subgroup moves the pseudo-I^2 down an interpretive category (to "30% to 60%: may represent moderate heterogeneity").
- If the pseudo-I^2 for Model 1b in the primary analysis was 80% ("50% to 90%: may represent substantial heterogeneity", but the pseudo-I^2 for the subgroup model for 'Exposed to human waste' (from subgroup analysis 1) was 40%, and the pseudo I^2 for the subgroup model for 'Untreated wastewater' (from subgroup analysis 2) was 25%, then certainty would be assessed for estimates for those exposed to human waste and those exposed to 'untreated wastewater' for each type of infection. This is because both subgroups move the pseudo-I^2 down an interpretive category (though to differing degrees; to "0% to 40%: might not be important" and "50% to 90%: may represent substantial heterogeneity", respectively).

Such splitting of the main outcomes into important subgroups before GRADE assessment and making the Summary of Findings table is recommended in the Cochrane Handbook (Schünemann et al., 2023).

Meta-regressions

Dose-response relationships

As the basis of our primary analysis, risk ratios comparing the group with the lowest exposure to polluted seawater and groups with higher exposures in a study provide a standardised metric of the association/effect across studies, without relying on studies measuring pollution exposure in the same way (Morgan et al., 2018). Nonetheless for studies that do measure pollution exposure using the same indicators, it may be possible to derive dose-response type relationships between the pollution indicator and risk ratios. This could be useful for increasing certainty in the effect (see 'Confidence in cumulative evidence'; Supplementary Information 10). It could also be useful for policymakers and other stakeholders - for example, to understand whether water quality guidelines are likely to prevent illness, which Wade et al., (2003) previously attempted to understand for gastrointestinal illness using synthesis methods).

In order to estimate these dose-response type effects, we will fit meta-regressions of the level of pollution to which exposure group was exposed (e.g. in faecal coliform CFU/ml water) against risk ratio. Following a similar logic to the subgroup analyses, such models will only be fit only for subsets of the data in which there remain at least two studies (the minimum needed for a 'meta-analysis') after splitting the data according to type of infection and pollution indicator (since we expect dose-response effects to differ according to both; Fleisher et al., 1996). The effect of dose will be fit as a linear term in the first instance, but if there is no significant linear effect we will attempt to fit dose as a restricted cubic spline (Desquilbet and Mariotti, 2010; Viechtbauer, 2023).

Again, like subgroup analyses, these analyses will primarily be exploratory, in recognition of the limited statistical power of subgroup analyses in meta-analyses (Cuijpers et al., 2021). However, if one of the dose-response models for an outcome indicates a significant dose-response relationship, then we will upgrade the certainty rating assigned to the pooled estimate in GRADE accordingly (Supplementary Information 10).

Meta-bias

Selective reporting within studies will be assessed at the outcome level (type of infection or important subgroup within type of infection) as part of risk of bias assessments (see 'Bias in individual studies'). Publication bias will be assessed for the quantitative synthesis, for the pooled estimates remaining after sensitivity and subgroup analyses (i.e. per type of infection, per type of infection after removal of high risk of bias studies, or per important subgroups within each type of infection).

To assess publication bias, we will first generate funnel plots of the individual estimates underpinning each pooled estimate as an initial method of exploring small study publication bias, for continuity with previous work. However, this method is now considered an unreliable way of testing for publication bias – for example because funnel plot interpretation is subjective, funnel plot shape is heavily influenced by the relatively arbitrary choice of method used to construct the plot, and funnel plot inspection is prone to mistakenly detecting bias where studies are heterogeneous (Sterne, Gavaghan and Egger, 2000; Tang and Liu, 2000; Terrin *et al.*, 2003). Therefore, we will formally assess small study publication bias for each pooled estimate using the more quantitative method of Egger's regression, correcting our estimates for this if detected, following the method of Nakagawa *et al.*, (2022). Results of this more quantitative analysis/correction will still be interpreted with caution, however, given the limitations of even this quantitative test (Sterne, Gavaghan and Egger, 2000).

Time-lag publication bias will be also assessed for each pooled estimate and accounted for if detected - again following the regression method of Nakagawa *et al.*, (2022). Time-lag publication bias tests test for a relationship between reported estimates of the association/effect and the year of the study/publication from whence they are derived. A negative relationship is theoretically expected to occur because more "exciting", big effects are more likely to be published earlier on in the timeline of research on a particular question. However, in this context declines or increases in reported estimates of association over time could also be indicative of changing wastewater management practices, awareness, or climate, for example. We will therefore also consider these alternative explanations when interpreting our time-lag publication bias test results and look for additional circumstantial evidence to support the idea of publication bias (see 'Confidence in cumulative evidence'; Supplementary Information 10).

P-values for publication bias tests for a particular outcome will be Bonferroni-corrected for multiple tests for publication bias (i.e. 2 tests – for one small study publication bias, and one for time-lag publication bias.

Confidence in cumulative evidence

Confidence in the cumulative evidence will be assessed separately for seawater, freshwater, and transitional water for each of the 6 infection types using the GRADE approach for systematic reviews, with the results presented in a Summary of Findings table (Balshem et al., 2011; GRADE Working Group, 2004, 2004; Schünemann et al., 2023). Specifically, we assess confidence in the pooled estimates from the quantitative synthesis for infections of the ear, eye, gastrointestinal tract, skin, respiratory tract, and urinary tract (though certainty will be assessed at the level of subgroup estimates if subgroup(s) appears to explain residual heterogeneity between types of infection; see Synthesis: Additional analyses: Subgroup analyses). GRADE assessment is limited to the quantitative synthesis because two of the five GRADE domains rely heavily on a pooled estimate (Inconsistency and Imprecision), and two others benefit from additional quantitative analyses to confirm their effects (Risk of Bias and Publication Bias).

To begin our implementation of the GRADE approach, the pooled estimate(s) for each type of infection will be assigned a 'high' initial level of certainty. It is anticipated that each body of evidence used in our review will be dominated by evidence from non-randomised studies, which is sometimes a reason given for assigning a lower initial level of certainty to bodies of evidence. However, we recognise that observational evidence is sometimes considered an especially helpful type of evidence in environmental health (Woodruff and Sutton, 2014). Combined with our assessments of risk of bias in non-randomised studies using the ROBINS-E tool - which can be used to downgrade the level of certainty appropriately (Schünemann et al., 2023) - this justifies our assignment of a 'high' initial level of certainty for evidence from both randomised and non-randomised studies.

Subsequently, we will consider whether to lower and, if necessary, raise the level of certainty in each pooled estimate. Following the Cochrane guidance (Schünemann et al., 2023) for the implementation of GRADE for systematic reviews, will first consider 5 downgrading domains (risk of bias, inconsistency, indirectness, imprecision, and publication bias), followed by 3 upgrading domains (large effects, dose-response, plausible confounding). We have developed pre-defined criteria to guide the implementation of downgrading/upgrading in each domain, which are based heavily on Cochrane and GRADE guidance (Supplementary Information 8). However, in developing these criteria we found that GRADE (perhaps purposefully) does not always make clear whether it is the estimate itself or its direction for which certainty is being assessed. At least in this context, we think that assessing certainty in the estimate itself is more useful, and therefore tailor certain domains accordingly using Cochrane and GRADE guidance where possible (namely imprecision, and plausible confounding). Readers can then make their own judgements about whether the direction of the effect is sufficiently convincing, based on the estimate information and GRADE judgement presented in the Summary of Findings table.

This GRADE-assessment process will be piloted by two reviewers before full implementation by assessing one of the 6 types of infection. To begin the piloting process, an initial meeting will be held between the two reviewers to clarify the use of both tools in the context of this body of literature, and clarify any initial uncertainties regarding the application of the tools. The two reviewers will then independently assess the body of evidence for the type of infection considered. Finally, the two reviewers will meet to discuss and resolve any discrepancies in certainty assessments for each domain, and clarify the guidance accordingly (with adjustments to the guidance reported in the final publication).

After the piloting process, GRADE assessments for all pooled estimates will be carried out in a double-blind manner by two reviewers working independently to populate the same Excel template (Microsoft Corporation, 2018). Any discrepancies between the two reviewers in GRADE rating for each pooled estimate will be resolved through mediation by a third reviewer. If a consensus cannot be reached, we will conservatively defer to the lowest certainty rating made by the three reviewers. A final 'certainty of the evidence' rating for each pooled estimate for each type of infection will be presented in the 'Summary of findings' table, alongside information on number of participants/studies, the estimate and its confidence intervals, and illustrative comparative risks (Schünemann et al., 2023).

Registration

This publication (DOI: XXXXXX) serves as the registration of this review, with the full review intended to be published in Environment International in line with the Registered Report format (Nosek and Lakens, 2014).

Data & code availability

Code for the analyses planned for this systematic review update is available via GitHub (https://github.com/befriendabacterium/REPOWI-SRMA_protocol/), with the version used for this manuscript permanently archived at Zenodo (https://doi.org/10.5281/10.5281/zenodo.12207065. The data used to develop this code and which is to be used alongside it is the meta-analysis dataset of the previous review (Leonard et al., 2018b), which was deposited at Open Research Exeter (https://doi.org/10.24378/exe.123).

Author contribution statement

NB This contribution statement follows the CRediT taxonomy (Brand et al., 2015) *and refers to the protocol, rather than the final review.* Matt Lloyd Jones is the guarantor of the final review.

Matt Lloyd Jones: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project administration; **Anne Francis Leonard:** Conceptualization, Methodology, Resources, Data Curation, Writing – Original Draft, Supervision, Project administration, Funding acquisition; **Alison Bethel:** Methodology, Software, Resources, Formal analysis, Data Curation, Writing – Original Draft; **Emma lamb:** Formal analysis, Investigation, Resources; **Will Gaze:** Conceptualization, Methodology, Project administration, Funding acquisition; Writing – Review & Editing; **Tim Taylor:** Conceptualization, Methodology, Project administration, Funding acquisition; Writing – Review & Editing; **Andrew Singer**: Conceptualization, Methodology, Writing – Review & Editing**; Obioha Ukoumunne:** Conceptualization, Methodology, Writing – Review & Editing; **Ruth Garside:** Conceptualization, Methodology, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

Support

This systematic review update is funded by the BlueAdapt project, which is part of the European climate change and health cluster and has received funding from the European Union's Horizon Europe research and innovation program and the UKRI/HM Government under Horizon Europe Grant Agreement No 101057764. The funders had no direct role in the development of this protocol.

Competing interests

The authors declare the following possible competing interests/conflicts of interest:

- AFL, WG, AS, OU and RG worked on the previous related systematic review (Leonard et al., 2018a)
- AFL, WG, AS, OU, and RG have previously been involved in primary studies in this field (Leonard et al., 2018c, 2020; Farrell et al., 2023), one of which is anticipated to be included in the review (Leonard et al., 2018c).
- MLJ, AFL, EL, and WG are currently involved in a research project with Surfers Against Sewage, which is anticipated to be included in the review (Leonard et al., 2022).
- MLJ, AFL, and WG have previously worked for/with/received research funding from the UK Environment Agency

Other

This systematic review protocol has been reported according to PRISMA-P guidelines (Shamseer et al., 2015; Table S4). The subsequent review will be reported according to PRISMA2020 guidelines (Page et al., 2021), unless more up to date PRISMA guidelines become available. Deviations from this protocol in the final review will be reported and justified via the tool of Willroth and Atherton (2024). A glossary of terms used in this review protocol is included in Supplementary information 11).

("What is the research evidence for antibiotic resistance exposure and transmission to humans from the environment? A systematic map protocol | Environmental Evidence," n.d.)

Bibliography

Adhikary, R.K., Mahfuj, M.S.-E., Starrs, D., Croke, B., Glass, K., Lal, A., 2022. Risk of Human Illness from Recreational Exposure to Microbial Pathogens in Freshwater Bodies: A Systematic Review. Expo Health 14, 325–343. https://doi.org/10.1007/s12403-021-00447-z

Allen, R.W., Barn, P.K., Lanphear, B.P., 2015. Randomized Controlled Trials in Environmental Health Research: Unethical or Underutilized? PLoS Med 12, e1001775. https://doi.org/10.1371/journal.pmed.1001775

Balshem, H., Helfand, M., Schünemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., Vist, G.E., Falck-Ytter, Y., Meerpohl, J., Norris, S., Guyatt, G.H., 2011. GRADE guidelines: 3. Rating the quality of evidence. Journal of Clinical Epidemiology 64, 401–406. https://doi.org/10.1016/j.jclinepi.2010.07.015

Boutron, I., Page, Higgins, Altman, Lundh, A., Hróbjartsson, A., 2024. Chapter 7: Considering bias and conflicts of interest among the included studies, in: Cochrane Handbook for Systematic Reviews of Interventions.

Brand, A., Allen, L., Altman, M., Hlava, M., Scott, J., 2015. Beyond authorship: attribution, contribution, collaboration, and credit. Learned Publishing 28, 151–155. https://doi.org/10.1087/20150211

Britton, E., Kindermann, G., Domegan, C., Carlin, C., 2020. Blue care: a systematic review of blue space interventions for health and wellbeing. Health Promot Int 35, 50–69. https://doi.org/10.1093/heapro/day103

Burke, J.F., Sussman, J.B., Kent, D.M., Hayward, R.A., 2015. Three simple rules to ensure reasonably credible subgroup analyses. BMJ 351, h5651. https://doi.org/10.1136/bmj.h5651

Burnett, N., Bolton, P., 2025. Sewage discharges.

Campbell, M., McKenzie, J.E., Sowden, A., Katikireddi, S.V., Brennan, S.E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., Thomson, H., 2020. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 368, l6890. https://doi.org/10.1136/bmj.l6890

Carroll, R., 2021. Lockdown has made us fall in love with the sea [WWW Document]. Irish Examiner. URL https://www.irishexaminer.com/lifestyle/healthandwellbeing/arid-40370333.html (accessed 4.8.24).

Cuijpers, P., Griffin, J.W., Furukawa, T.A., 2021. The lack of statistical power of subgroup analyses in meta-analyses: a cautionary note. Epidemiol Psychiatr Sci 30, e78. https://doi.org/10.1017/S2045796021000664

Cummings, P., 2009. The relative merits of risk ratios and odds ratios. Arch Pediatr Adolesc Med 163, 438–445. https://doi.org/10.1001/archpediatrics.2009.31

Dani, A., 2014. Colonization and infection. Cent European J Urol 67, 86–87. https://doi.org/10.5173/ceju.2014.01.art19

Deeks, J., Higgins, J.P., Altman, D.P., Cochrane Statistical Methods Group, 2024. Chapter 10: Analysing data and undertaking meta-analyses, in: Cochrane Handbook for Systematic Reviews of Interventions.

DEFRA, 2024. New multi-million pound action plan launched to protect the River Wye [WWW Document]. GOV.UK. URL https://www.gov.uk/government/news/new-multi-million-pound-action-plan-launched-to-protect-the-river-wye (accessed 1.21.25).

Desquilbet, L., Mariotti, F., 2010. Dose-response analyses using restricted cubic spline functions in public health research. Stat Med 29, 1037–1057. https://doi.org/10.1002/sim.3841

Environment Agency, 2024a. Environment Agency publishes storm overflow spill data for 2023 [WWW Document]. 27/03/24. URL https://www.gov.uk/government/news/environment-agency-publishes-storm-overflow-spill-data-for-2023 (accessed 4.8.24).

Environment Agency, 2024b. 2023 Event Duration Monitoring Data Publication – Creating a better place.

Environment Agency, 2024c. Environment Agency explains that pollution in Windermere is reducing but more needs to be done to tackle all sources of pollution, not just sewage – Creating a better place [WWW Document]. URL https://environmentagency.blog.gov.uk/2024/05/24/environment-agency-explains-that-pollution-in-windermere-is-reducing-but-more-needs-to-be-done-to-tackle-all-sources-of-pollution-not-just-sewage/ (accessed 1.21.25).

Environment Agency, 2023a. Storm overflow spill data shows performance is totally unacceptable – Creating a better place [WWW Document]. URL https://environmentagency.blog.gov.uk/2023/03/31/storm-overflow-spill-data-shows-performance-is-totally-unacceptable/ (accessed 4.8.24).

Environment Agency, 2023b. New data identifies amount of pollution from different sources in Windermere – Creating a better place [WWW Document]. GOV.UK. URL https://environmentagency.blog.gov.uk/2023/12/20/new-data-identifies-amount-of-pollution-from-different-sources-in-windermere/ (accessed 1.24.25).

Farrell, M.L., Chueiri, A., O'Connor, L., Duane, S., Maguire, M., Miliotis, G., Cormican, M., Hooban, B., Leonard, A., Gaze, W.H., Devane, G., Tuohy, A., Burke, L.P., Morris, D., 2023. Assessing the impact of recreational water use on carriage of antimicrobial resistant organisms. Science of The Total Environment 888, 164201. https://doi.org/10.1016/j.scitotenv.2023.164201

Fleisher, J.M., Fleming, L.E., Solo-Gabriele, H.M., Kish, J.K., Sinigalliano, C.D., Plano, L., Elmir, S.M., Wang, J.D., Withum, K., Shibata, T., Gidley, M.L., Abdelzaher, A., He, G., Ortega, C., Zhu, X., Wright, M., Hollenbeck, J., Backer, L.C., 2010. The BEACHES Study: health effects and exposures from non-point source microbial contaminants in subtropical recreational marine waters. International Journal of Epidemiology 39, 1291–1298. https://doi.org/10.1093/ije/dyq084

Fleisher, J.M., Kay, D., 2006. Risk perception bias, self-reporting of illness, and the validity of reported results in an epidemiologic study of recreational water associated illnesses. Mar Pollut Bull 52, 264–268. https://doi.org/10.1016/j.marpolbul.2005.08.019

Fleisher, J.M., Kay, D., Salmon, R.L., Jones, F., Wyer, M., Godfree, A.F., 1996. Marine waters contaminated with domestic sewage: Nonenteric illnesses associated with bather exposure in the United Kingdom. American Journal of Public Health 86, 1228–1234. https://doi.org/10.2105/AJPH.86.9.1228

Frampton, G., Whaley, P., Bennett, M., Bilotta, G., Dorne, J.-L.C.M., Eales, J., James, K., Kohl, C., Land, M., Livoreil, B., Makowski, D., Muchiri, E., Petrokofsky, G., Randall, N., Schofield, K., 2022. Principles and framework for assessing the risk of bias for studies included in comparative quantitative environmental systematic reviews. Environmental Evidence 11, 12. https://doi.org/10.1186/s13750-022-00264-0

Freiman, A., Barankin, B., Elpern, D.J., 2004. Sports dermatology part 2: swimming and other aquatic sports. CMAJ 171, 1339–1341. https://doi.org/10.1503/cmaj.1040892

Gasperi, J., Gromaire, M.C., Kafi, M., Moilleron, R., Chebbo, G., 2010. Contributions of wastewater, runoff and sewer deposit erosion to wet weather pollutant loads in combined sewer systems. Water Research 44, 5875–5886. https://doi.org/10.1016/j.watres.2010.07.008

Gogien, F., Dechesne, M., Martinerie, R., Lipeme Kouyi, G., 2023. Assessing the impact of climate change on Combined Sewer Overflows based on small time step future rainfall timeseries and long-term continuous sewer network modelling. Water Research 230, 119504. https://doi.org/10.1016/j.watres.2022.119504

GRADE Working Group, 2004. Grading quality of evidence and strength of recommendations. BMJ 328, 1490. https://doi.org/10.1136/bmj.328.7454.1490

Grainger, M., Stewart, G., Haddaway, N.R., 2022. Why "vote-counting" is never acceptable in evidence synthesis. https://doi.org/10.31219/osf.io/c49uh

Grant, R.L., 2014a. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. BMJ 348, f7450. https://doi.org/10.1136/bmj.f7450

Grant, R.L., 2014b. Converting Between Probabilities, Odds (Ratios), and Risk Ratios [vignette for "effectsize" R package].

Gusenbauer, M., 2019. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. Scientometrics 118, 177–214. https://doi.org/10.1007/s11192-018-2958-5

Gusenbauer, M., Haddaway, N.R., 2020. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Research Synthesis Methods 11, 181–217. https://doi.org/10.1002/jrsm.1378

Haddaway, N.R., Collins, A.M., Coughlin, D., Kirk, S., 2015. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. PLoS One 10, e0138237. https://doi.org/10.1371/journal.pone.0138237

Haddaway, N.R., Grainger, M.J., Gray, C.T., 2022. Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. Research Synthesis Methods 13, 533–545. https://doi.org/10.1002/jrsm.1563

Haddaway, N.R., Macura, B., Whaley, P., Pullin, A.S., 2018. ROSES RepOrting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. Environmental Evidence 7, 7. https://doi.org/10.1186/s13750-018-0121-7

Hammond, P., Suttie, M., Lewis, V.T., Smith, A.P., Singer, A.C., 2021. Detection of untreated sewage discharges to watercourses using machine learning. npj Clean Water 4, 1–10. https://doi.org/10.1038/s41545-021-00108-3

Hannes, K., Lockwood, C., Pearson, A., 2010. A Comparative Analysis of Three Online Appraisal Instruments' Ability to Assess Validity in Qualitative Research. Qual Health Res 20, 1736–1743. https://doi.org/10.1177/1049732310378656

Harzing, A.-W., 2007. Publish or Perish.

Hensyl, B.F., Borhani, S., Payab, A.H., Montalto, F., 2024. Opportunities for Leveraging Existing Hydrologic and Hydraulic Models Developed for Water Quantity Management to Mitigate Flooding Due to Extreme Precipitation. JWMM. https://doi.org/10.14796/JWMM.C516

Hernán, M.A., Hernández-Díaz, S., 2012. Beyond the intention to treat in comparative effectiveness research. Clin Trials 9, 48–55. https://doi.org/10.1177/1740774511420743

Higgins, J.P., Li, T., Deeks, J.J., 2023. Chapter 6: Choosing effect measures and computing estimates of effect.

Higgins, J.P., Savović, J., Page, M.J., Elbers, R.G., Sterne, J.A., 2024. Chapter 8: Assessing risk of bias in a randomized trial, in: Cochrane Handbook for Systematic Reviews of Interventions.

Higgins, J.P., Savović, J., Page, M.J., Sterne, onathan A., 2019. Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) [guidance notes for parallel trials].

Higgins, J.P.T., Morgan, R.L., Rooney, A.A., Taylor, K.W., Thayer, K.A., Silva, R.A., Lemeris, C., Akl, E.A., Bateson, T.F., Berkman, N.D., Glenn, B.S., Hróbjartsson, A., LaKind, J.S., McAleenan, A., Meerpohl, J.J., Nachman, R.M., Obbagy, J.E., O'Connor, A., Radke, E.G., Savović, J., Schünemann, H.J., Shea, B., Tilling, K., Verbeek, J., Viswanathan, M., Sterne, J.A.C., 2024. A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). Environment International 186, 108602. https://doi.org/10.1016/j.envint.2024.108602

Hill, A., Mirchandani, M., Pilkington, V., 2022. Ivermectin for COVID-19: Addressing Potential Bias and Medical Fraud. Open Forum Infectious Diseases 9, ofab645. https://doi.org/10.1093/ofid/ofab645

Honda, R., Tachi, C., Yasuda, K., Hirata, T., Noguchi, M., Hara-Yamamura, H., Yamamoto-Ikemoto, R., Watanabe, T., 2020. Estimated discharge of antibiotic-resistant bacteria from combined sewer overflows of urban sewage system. npj Clean Water 3, 1–7. https://doi.org/10.1038/s41545-020-0059-5

ICGL, 2014. Third International Conference on Grey Literature in 1997 (ICGL Luxembourg definition, 1997 - Expanded in New York, 2004).

IntHout, J., Ioannidis, J.P.A., Rovers, M.M., Goeman, J.J., 2016. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 6, e010247. https://doi.org/10.1136/bmjopen-2015-010247

Izcovich, A., Peiris, S., Ragusa, M., Tortosa, F., Rada, G., Aldighieri, S., Reveiz, L., 2022a. Bias as a source of inconsistency in ivermectin trials for COVID-19: A systematic review. Ivermectin's suggested benefits are mainly based on potentially biased results. Journal of Clinical Epidemiology 144, 43–55. https://doi.org/10.1016/j.jclinepi.2021.12.018

Izcovich, A., Peiris, S., Ragusa, M., Tortosa, F., Rada, G., Aldighieri, S., Reveiz, L., 2022b. Bias as a source of inconsistency in ivermectin trials for COVID-19: A systematic review. Ivermectin's suggested benefits are mainly based on potentially biased results. Journal of Clinical Epidemiology 144, 43–55. https://doi.org/10.1016/j.jclinepi.2021.12.018

Jackson, D., White, I.R., Riley, R.D., 2012. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. Stat Med 31, 3805–3820. https://doi.org/10.1002/sim.5453

Kay, D., Jones, F., Wyer, M.D., Fleisher, J.M., Salmon, R.L., Godfree, A.F., Zelenauch-Jacquotte, A., Shore, R., 1994. Predicting likelihood of gastroenteritis from sea bathing: results from randomised exposure. The Lancet 344, 905–909. https://doi.org/10.1016/S0140-6736(94)92267-5

Kroon, D.F., Lawson, M.L., Derkay, C.S., Hoffmann, K., McCook, J., 2002. Surfer's Ear: External Auditory Exostoses are More Prevalent in Cold Water Surfers. Otolaryngology–Head and Neck Surgery 126, 499–504. https://doi.org/10.1067/mhn.2002.124474

Lajeunesse, M.J., 2011. On the meta-analysis of response ratios for studies with correlated and multi-group designs. Ecology 92, 2049–2055. https://doi.org/10.1890/11-0423.1

Lakens, D., Mesquida, C., Rasti, S., Ditroilo, M., 2024. The benefits of preregistration and Registered Reports. Evidence-Based Toxicology 2, 2376046. https://doi.org/10.1080/2833373X.2024.2376046

Laville, S., Horton, H., Clark, A., correspondent, environment, 2024. Water companies in England face outrage over record sewage discharges. The Guardian.

Lemmin-Woolfrey, U., 2021. Why wild swimming is Britain's new craze [WWW Document]. URL https://www.bbc.com/travel/article/20210603-why-wild-swimming-is-britains-new-craze (accessed 4.8.24).

Leonard, A., Duane, S., Gaze, W., Taylor, T., 2022. Evaluating a bathing water quality app. BlueAdapt. URL https://blueadapt.eu/project/evaluating-a-sea-pollution-alert-app/ (accessed 10.7.24).

Leonard, A.F.C., Garside, R., Ukoumunne, O.C., Gaze, W.H., 2020. A cross-sectional study on the prevalence of illness in coastal bathers compared to non-bathers in England and Wales: Findings from the Beach User Health Survey. Water Research 176, 115700. https://doi.org/10.1016/j.watres.2020.115700

Leonard, A.F.C., Singer, A., Ukoumunne, O.C., Gaze, W.H., Garside, R., 2018a. Is it safe to go back into the water? A systematic review and meta-analysis of the risk of acquiring infections from recreational exposure to seawater. Int J Epidemiol 47, 572–586. https://doi.org/10.1093/ije/dyx281

Leonard, A.F.C., Singer, A., Ukoumunne, O.C., Gaze, W.H., Garside, R., 2018b. Is it safe to go back into the water? A systematic review and meta-analysis of the risk of acquiring infections from recreational exposure to seawater (dataset). https://doi.org/10.24378/exe.123

Leonard, A.F.C., Zhang, L., Balfour, A.J., Garside, R., Hawkey, P.M., Murray, A.K., Ukoumunne, O.C., Gaze, W.H., 2018c. Exposure to and colonisation by antibiotic-resistant *E. coli* in UK coastal water users: Environmental surveillance, exposure assessment, and epidemiological study (Beach Bum Survey). Environment International 114, 326–333. https://doi.org/10.1016/j.envint.2017.11.003

Malik, O., Jao, P., Hsu, A., 2014. Assessing global water quality: The data challenge [WWW Document]. Global Water Forum. URL https://www.globalwaterforum.org/2014/12/09/assessing-global-water-quality-the-data-challenge/ (accessed 1.24.25).

Mannocci, A., La Torre, G., Spagnoli, A., Solimini, A.G., Palazzo, C., De Giusti, M., 2016. Is swimming in recreational water associated with the occurrence of respiratory illness? A systematic review and meta-analysis. Journal of Water and Health 14, 590–599. https://doi.org/10.2166/wh.2016.266

Mathers, M., 2024. Raw sewage spills into rivers and seas in England reached record high last year. The Independent.

McGuinness, L.A., Higgins, J.P.T., 2021. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. Research Synthesis Methods 12, 55–61. https://doi.org/10.1002/jrsm.1411

McKenzie, J.E., Brennan, S.E., Cochrane GRADEing Methods Group, Cochrane Statistical Methods Group, 2023. Chapter 12: Synthesizing and presenting findings using other methods.

Mei, F., Chen, F., Hu, K., Gao, Q., Zhao, L., Shang, Y., Zhao, B., Ma, B., 2022. Registration and Reporting Quality of Systematic Reviews on Surgical Intervention: A Meta-epidemiological Study. J Surg Res 277, 200–210. https://doi.org/10.1016/j.jss.2022.04.026

Microsoft Corporation, 2018. Microsoft excel.

Moher, D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D.F., 1999. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. The Lancet 354, 1896–1900. https://doi.org/10.1016/S0140-6736(99)04149-5

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339, b2535. https://doi.org/10.1136/bmj.b2535

Moran, N.P., Sánchez-Tójar, A., Schielzeth, H., Reinhold, K., 2021. Poor nutritional condition promotes high-risk behaviours: a systematic review and meta-analysis. Biological Reviews 96, 269–288. https://doi.org/10.1111/brv.12655

Morgan, R.L., Whaley, P., Thayer, K.A., Schünemann, H.J., 2018. Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. Environment International 121, 1027–1031. https://doi.org/10.1016/j.envint.2018.07.015

Nakagawa, S., Lagisz, M., Jennions, M.D., Koricheva, J., Noble, D.W.A., Parker, T.H., Sánchez-Tójar, A., Yang, Y., O'Dea, R.E., 2022. Methods for testing publication bias in ecological and evolutionary meta-analyses. Methods in Ecology and Evolution 13, 4–21. https://doi.org/10.1111/2041-210X.13724

Nakagawa, S., Lagisz, M., O'Dea, R.E., Rutkowska, J., Yang, Y., Noble, D.W.A., Senior, A.M., 2021. The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond. Research Synthesis Methods 12, 4–12. https://doi.org/10.1002/jrsm.1424

Nakagawa, S., Yang, Y., Macartney, E.L., Spake, R., 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. https://doi.org/10.1186/s13750-023-00301-6

National Engineering Policy Centre Working Group, 2024. Testing the waters report: reducing health risks from water pollution. National Engineering Policy Centre.

Noble, D.W.A., Lagisz, M., O'dea, R.E., Nakagawa, S., 2017. Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. Molecular Ecology 26, 2410–2425. https://doi.org/10.1111/mec.14031

Nosek, B.A., Lakens, D., 2014. Registered reports: A method to increase the credibility of published results. Social Psychology 45, 137–141. https://doi.org/10.1027/1864-9335/a000192

O'Dea, R.E., Lagisz, M., Hendry, A.P., Nakagawa, S., 2019. Developmental temperature affects phenotypic means and variability: A meta-analysis of fish data. Fish and Fisheries 20, 1005–1022. https://doi.org/10.1111/faf.12394

OFWAT, 2008. Preparing for the future – Ofwat's climate change policy statement. OFWAT, Birmingham.

Outdoor Swimmer, 2021. Outdoor Swimmer Trends Report 2021.

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan—a web and mobile app for systematic reviews. Systematic Reviews 5, 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71. https://doi.org/10.1136/bmj.n71

Park, A., 2024. How Dangerous Is the Seine for Athletes? [WWW Document]. TIME. URL https://time.com/6999263/seine-health-risks-paris-olympics/ (accessed 1.21.25).

Pistocchi, A., 2020. A preliminary pan-European assessment of pollution loads from urban runoff. Environmental Research 182, 109129. https://doi.org/10.1016/j.envres.2020.109129

Reeves, B.C., Deeks, J.J., Higgins, J.P., Shea, B., Tugwell, P., Wells, G.A., Cochrane Non-Randomized Studies of Interventions Methods Group, 2023. Chapter 24: Including non-randomized studies on intervention effects.

Rethlefsen, M.L., Farrell, A.M., Trzasko, L.C.O., Brigham, T.J., 2015. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. Journal of Clinical Epidemiology 68, 617–626. https://doi.org/10.1016/j.jclinepi.2014.11.025

Rethlefsen, M.L., Page, M.J., 2022. PRISMA 2020 and PRISMA-S: common questions on tracking records and the flow diagram. J Med Libr Assoc 110, 253–257. https://doi.org/10.5195/jmla.2022.1449

Reynolds, G., Amenabar, T., Beesch, C., Schroer, B., Tambe, A., Sima, R., 2024. Did swimming in the Seine make athletes sick? Here's what we know. Washington Post.

Rubio-Aparicio, M., López-López, J.A., Viechtbauer, W., Marín-Martínez, F., Botella, J., Sánchez-Meca, J., 2020. Testing Categorical Moderators in Mixed-Effects Meta-analysis in the Presence of Heteroscedasticity. The Journal of Experimental Education 88, 288–310. https://doi.org/10.1080/00220973.2018.1561404

Rubio-Aparicio, M., Sánchez-Meca, J., López-López, J.A., Botella, J., Marín-Martínez, F., 2017. Analysis of categorical moderators in mixed-effects meta-analysis: Consequences of using pooled versus separate estimates of the residual between-studies variances. British Journal of Mathematical and Statistical Psychology 70, 439–456. https://doi.org/10.1111/bmsp.12092

Russo, G.S., Eftim, S.E., Goldstone, A.E., Dufour, A.P., Nappier, S.P., Wade, T.J., 2020. Evaluating health risks associated with exposure to ambient surface waters during recreational activities: A systematic review and meta-analysis. Water Research 176, 115729. https://doi.org/10.1016/j.watres.2020.115729

Sasse, T., 2024. Why so many Britons have taken to stand-up paddleboarding. The Economist.

Schünemann, H., Brożek, J., Guyatt, G., Oxman, A., Akl, E., Mustafa, R., Wiercioch, N.S.W., 2013. GRADE handbook for grading quality of evidence and strength of recommendations, October 2013. ed.

Schünemann, H.J., Higgins, J.P.T., Vist, G.E., Glasziou, P., Akl, E.A., Skoetz, N., Guyatt, G.H., Cochrane GRADEing Methods Group, Cochrane Statistical Methods Group, 2023. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis

protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 349, g7647. https://doi.org/10.1136/bmj.g7647

S

Shojania, K.G., Sampson, M., Ansari, M.T., Ji, J., Doucette, S., Moher, D., 2007. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med 147, 224–233. https://doi.org/10.7326/0003-4819-147-4-200708210-00179

Smith, R., Ormerod, J.O.M., Sabharwal, N., Kipps, C., 2018. Swimming-induced pulmonary edema: current perspectives. Open Access J Sports Med 9, 131–137. https://doi.org/10.2147/OAJSM.S140028

Sport England, 2024. Active Lives Adult Survey November 2022-23 Report.

Stang, A., 2010. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol 25, 603–605. https://doi.org/10.1007/s10654-010-9491-z

Stang, A., Jonas, S., Poole, C., 2018. Case study in major quotation errors: a critical commentary on the Newcastle–Ottawa scale. Eur J Epidemiol 33, 1025–1031. https://doi.org/10.1007/s10654-018-0443-3

Sterne, J.A., Hernán, M.A., Reeves, B.C., Savović, J., Berkman, N.D., Viswanathan, M., Henry, D., Altman, D.G., Ansari, M.T., Boutron, I., Carpenter, J.R., Chan, A.-W., Churchill, R., Deeks, J.J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y.K., Pigott, T.D., Ramsay, C.R., Regidor, D., Rothstein, H.R., Sandhu, L., Santaguida, P.L., Schünemann, H.J., Shea, B., Shrier, I., Tugwell, P., Turner, L., Valentine, J.C., Waddington, H., Waters, E., Wells, G.A., Whiting, P.F., Higgins, J.P., 2016. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 355, i4919. https://doi.org/10.1136/bmj.i4919

Sterne, J.A.C., Gavaghan, D., Egger, M., 2000. Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. Journal of Clinical Epidemiology 53, 1119–1129. https://doi.org/10.1016/S0895-4356(00)00242-0

Sterne, J.A.C., Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., Cates, C.J., Cheng, H.-Y., Corbett, M.S., Eldridge, S.M., Emberson, J.R., Hernán, M.A., Hopewell, S., Hróbjartsson, A., Junqueira, D.R., Jüni, P., Kirkham, J.J., Lasserson, T., Li, T., McAleenan, A., Reeves, B.C., Shepperd, S., Shrier, I., Stewart, L.A., Tilling, K., White, I.R., Whiting, P.F., Higgins, J.P.T., 2019. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 366, l4898. https://doi.org/10.1136/bmj.l4898

Surf Industry Members Association, 2024. Diversity in Surfing.

Tang, J.-L., Liu, J.L., 2000. Misleading funnel plot for detection of bias in meta-analysis. Journal of Clinical Epidemiology 53, 477–484. https://doi.org/10.1016/S0895-4356(99)00204-8

Terrin, N., Schmid, C.H., Lau, J., Olkin, I., 2003. Adjusting for publication bias in the presence of heterogeneity. Statistics in Medicine 22, 2113–2126. https://doi.org/10.1002/sim.1461

The Big Blue Swim, 2020. OPEN WATER SWIMMING INCREASES DRAMATICALLY DURING COVID-19 PANDEMIC [WWW Document]. The Big Blue Swim. URL https://thebigblueswim.com/blog/post/2020-12-23-open-water-swimming-increases-dramatically-during-covid-19-pandemic (accessed 4.8.24).

The EndNote Team, 2024. EndNote.

The Nursery Research and Planning, 2022. Watersports Participation Survey.

The Nursery Research and Planning, 2021. Watersports Participation Survey, Watersports Participation Survey.

Tien, C., 2024. Here's What a Microbiologist Thinks of the Seine Olympic Swimming Situation [WWW Document]. SELF. URL https://www.self.com/story/seine-swimming-water-quality-olympics (accessed 1.21.25).

Tipper, H.J., Stanton, I.C., Payne, R.A., Read, D.S., Singer, A.C., 2024. Do storm overflows influence AMR in the environment and is this relevant to human health? A UK perspective on a global issue. Water Research 260, 121952. https://doi.org/10.1016/j.watres.2024.121952

UNCTAD, 2024. Classifications | UNCTAD Data Hub [WWW Document]. Classifications. URL https://unctadstat.unctad.org/EN/Classifications.html (accessed 2.6.25).

UNCTAD, 2023. UNCTADstat [WWW Document]. From the Handbook of Statistics to Data Insights. URL https://unctadstat.unctad.org/EN/FromHbsToDataInsights.html (accessed 2.6.25).

UNCTAD, 2022. UNCTAD Handbook of Statistics 2022. Handbook of Statistics.

Viechtbauer, W., 2024. Meta-Analysis via Multivariate/Multilevel Linear (Mixed-Effects) Models — rma.mv [WWW Document]. URL https://wviechtb.github.io/metafor/reference/rma.mv.html (accessed 2.28.24).

Viechtbauer, W., 2023. Modeling Non-Linear Associations in Meta-Regression [The metafor Package] [WWW Document]. URL https://www.metafor-project.org/doku.php/tips:non_linear_meta_regression (accessed 2.28.24).

Viechtbauer, W., 2022. Bootstrapping rma.mv() model to get CI for pseudo R^2 statistic [WWW Document]. Gist. URL https://gist.github.com/wviechtb/6fbfca40483cb9744384ab4572639169 (accessed 3.11.24).

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software 36, 1–48. https://doi.org/10.18637/jss.v036.i03

von Hippel, P.T., 2015. The heterogeneity statistic I2 can be biased in small meta-analyses. BMC Medical Research Methodology 15, 35. https://doi.org/10.1186/s12874-015-0024-z

Wade, T.J., Pai, N., Eisenberg, J.N.S., Colford, J.M., 2003. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environ Health Perspect 111, 1102–1109. https://doi.org/10.1289/ehp.6241

Whaley, P., Letcher, R.J., Covaci, A., Alcock, R., 2016. Raising the standard of systematic reviews published in Environment International. Environment International 97, 274–276. https://doi.org/10.1016/j.envint.2016.08.007

Whaley, P., Roth, N., 2022. How we promote rigour in systematic reviews and evidence maps at Environment International. Environment International 170, 107543. https://doi.org/10.1016/j.envint.2022.107543

What is the research evidence for antibiotic resistance exposure and transmission to humans from the environment? A systematic map protocol | Environmental Evidence [WWW Document], n.d. URL https://link.springer.com/article/10.1186/s13750-020-00197-6 (accessed 10.24.24).

Wiedenmann, A., Krüger, P., Dietz, K., López-Pila, J.M., Szewzyk, R., Botzenhart, K., 2006. A randomized controlled trial assessing infectious disease risks from bathing in fresh recreational waters in relation to the concentration of Escherichia coli, intestinal

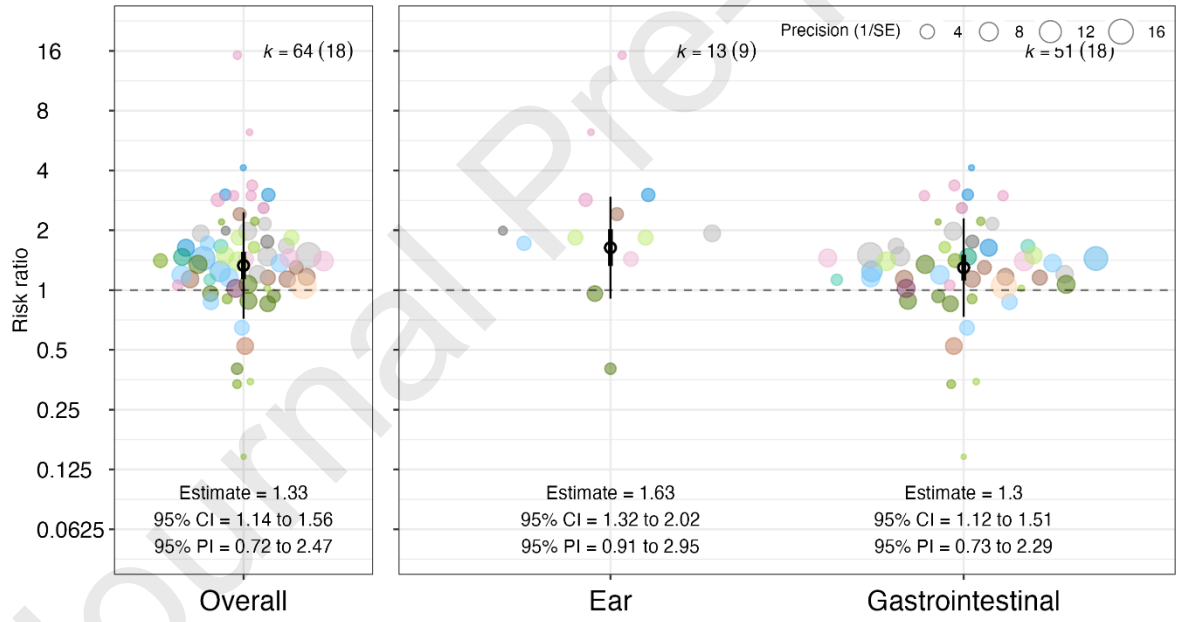
enterococci, Clostridium perfringens, and somatic coliphages. Environ Health Perspect 114, 228–236. https://doi.org/10.1289/ehp.8115

Wikoff, D., Lewis, R.J., Erraguntla, N., Franzen, A., Foreman, J., 2020. Facilitation of risk assessment with evidence-based methods – A framework for use of systematic mapping and systematic reviews in determining hazard, developing toxicity values, and characterizing uncertainty. Regulatory Toxicology and Pharmacology 118, 104790. https://doi.org/10.1016/j.yrtph.2020.104790

Willroth, E.C., Atherton, O.E., 2024. Best Laid Plans: A Guide to Reporting Preregistration Deviations. Advances in Methods and Practices in Psychological Science 7, 25152459231213802. https://doi.org/10.1177/25152459231213802

Woodruff, T.J., Sutton, P., 2014. The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes. Environmental Health Perspectives 122, 1007–1014. https://doi.org/10.1289/ehp.1307175

Yau, V., Wade, T.J., De Wilde, C.K., Colford, J.M., 2009. Skin-related symptoms following exposure to recreational water: a systematic review and meta-analysis. Water Expo. Health 1, 79–103. https://doi.org/10.1007/s12403-009-0012-9



Author contribution statement

NB This contribution statement follows the CRediT taxonomy (Brand et al., 2015) *and refers to the protocol, rather than the final review.* Matt Lloyd Jones is the guarantor of the final review.

Matt Lloyd Jones: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project administration; **Anne Francis Leonard:** Conceptualization,

Methodology, Resources, Data Curation, Writing – Original Draft, Supervision, Project administration, Funding acquisition; **Alison Bethel:** Methodology, Software, Resources, Formal analysis, Data Curation, Writing – Original Draft; **Emma lamb:** Formal analysis, Investigation, Resources; **Will Gaze:** Conceptualization, Methodology, Project administration, Funding acquisition; Writing – Review & Editing; **Tim Taylor:** Conceptualization, Methodology, Project administration, Funding acquisition; Writing – Review & Editing; **Andrew Singer**: Conceptualization, Methodology, Writing – Review & Editing**; Obioha Ukoumunne:** Conceptualization, Methodology, Writing – Review & Editing; **Ruth Garside:** Conceptualization, Methodology, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

MLJ, AFL, WG, TT and RG report that financial support was provided by the European Union's Horizon Europe research and innovation program and the UKRI/HM Government under Horizon Europe grant agreement No 101057764. Some co-authors have previously collaborated (AFL, OCU, RG, WG) and/or currently collaborate with Surfers Against Sewage (MLJ, AFL, RG, WG) on research projects related to the health risks associated with exposure to potentially-polluted seawater. Similarly, some authors have previously worked for or collaborated with, and/or currently collaborate with UK Environment Agency (MLJ, AFL, ACS, WG) on research projects related to antimicrobial resistance in the environment. Other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.