Blue-Green Systems

IWA PUBLISHING

Blue-Green Systems Vol 6 No 1, 133 doi: 10.2166/bgs.2024.049

A comparative analysis of international guidelines for green infrastructure performance assessment

Bardia Roghani [a,*], Mahdi Bahrami [a], Franz Tscheikner-Gratl [a], Frédéric Cherqui [b], Tone Merete Muthanna [a] and Marius Møller Rokstad [a]

[a] Norwegian University of Science and Technology (NTNU), Department of Civil and Environmental Engineering, Water and wastewater engineering (VA) group, Postbox 7491, Trondheim, Norway
[b] Univ. Lyon, INSA Lyon, Université Claude Bernard Lyon 1, DEEP, EA 7429, F-69621, F-69622, Villeurbanne Cedex, France
*Corresponding author. E-mail: bardia.roghani@ntnu.no

BR, 0000-0001-5435-6748; MB, 0000-0003-1399-3008; FT, 0000-0002-2545-6683; FC, 0000-0001-8836-3970; TMM, 0000-0002-4438-2202; MMR, 0000-0003-4650-3554

ABSTRACT

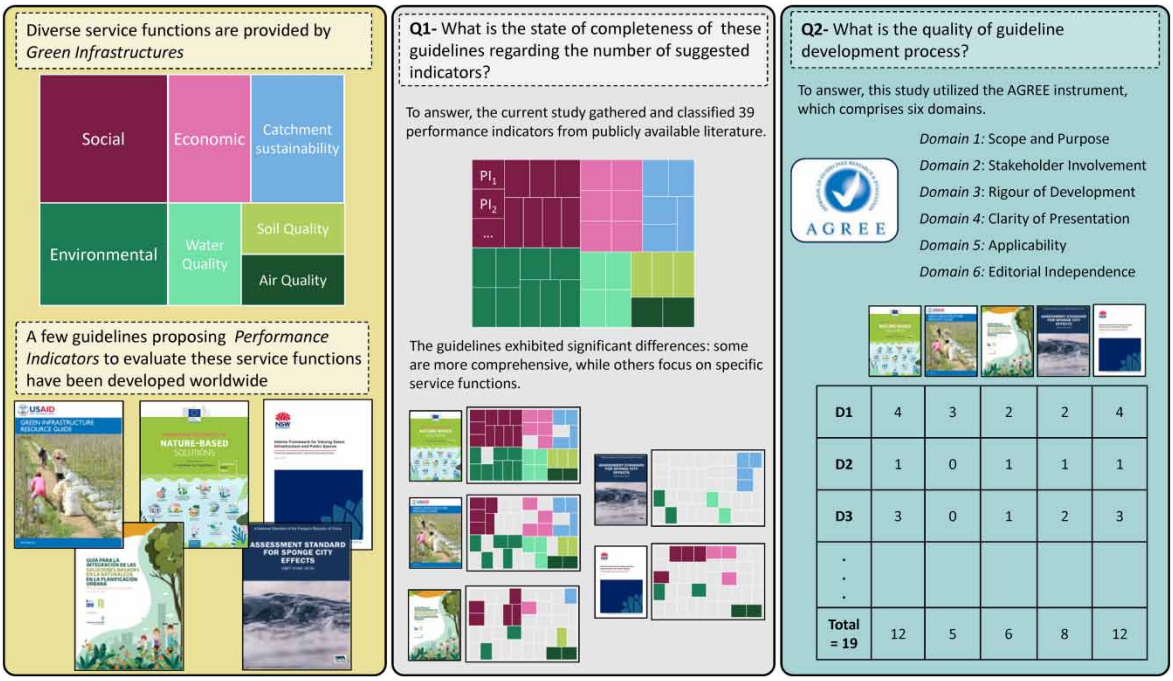
Natural or semi-natural elements, referred to by various terms such as green infrastructure (GI), are increasingly adopted as a win-win solution to develop water-wise, climate-resilient, and sustainable societies. Accordingly, a comprehensive performance assessment of GIs is necessary for their management, making them more appealing to multiple sectors, and improving funding prospects. Several guidelines proposing performance and impact indicators have been developed worldwide recently. To evaluate their applicability, this study critically examines some of the most well-known of them from various parts of the world that deal with the evaluation of different functional aspects of GIs. Findings show considerable differences. The European guideline is the most comprehensive one considering the number of addressed performance indicators. In contrast, the Chinese standard mostly focuses on water quality/quantity performance. Moreover, the degree of quality of the guidelines is evaluated using a thorough set of quality measures that consists of 19 criteria. By addressing 12 out of the 19 quality criteria, the European and state of New South Wales standards encompassed more guideline compilation requirements than the others. However, the way in which assessed performance indicators should be interpreted is one instance where a gap in the present standards is believed to be especially significant.

Key words: asset management, decision-making, green infrastructure, guideline, monitoring, multi-functionality, performance indicators

HIGHLIGHTS

- Guidelines have been recently developed to aid in selecting performance indicators for green infrastructure.
- This study critically examines multiple guidelines by setting quality measures.
- Considerable differences exist between the studied international performance assessment guidelines.
- Interpretation of performance indicators remains a weak point that can affect decision-making.

GRAPHICAL ABSTRACT



1. INTRODUCTION

During the 1980s, urban planning and development began considering natural water and nutrient cycles, with water flows managed in a way to replicate the original ecosystem (Radcliffe 2019). In North America, these philosophies were developed as *Best Management Practice* (*BMP*), then as *Low Impact Development (LID)*, and *Green Infrastructure (GI)*. While BMPs started with a focus on stormwater pollution prevention activities, the original intent of LID was to achieve natural catchment hydrology, and GIs covered concepts beyond stormwater management. In Britain, the term *Sustainable Urban Drainage Systems* (*SuDS*) has been adopted. Similar policies were developed in Australia under the frame of the *Water Sensitive Urban Design* (*WSUD*) philosophy aiming to minimize the hydrological impacts of urban development (Fletcher *et al.* 2015). Taking inspiration from concepts used worldwide, the *Sponge City* concept was developed in China in 2012 to redefine the relationship between people, water, and cities (Griffiths *et al.* 2020). In 2016, the IWA proposed the Principles for Water Wise Cities to assist urban leaders to develop and implement their vision for sustainable urban water, and resilient planning and design in their cities. Furthermore, the European Commission (2015) adopted the concept of nature-based solutions (*NBS*) as measures inspired by nature, which are cost-effective, simultaneously provide environmental, social, and economic benefits, and help build water-wise societies (Raymond *et al.* 2017; Ugarelli *et al.* 2021). It is evident that while they all refer to the same concept, different countries have adopted different names for it. This is due to the informal evolution of terminology used in urban drainage management, which is shaped by local and regional perspectives, understandings, and contexts (Fletcher *et al.* 2015). In the present study, inspired by Fletcher *et al.* (2015), the term GI will be used to encompass all expected service functions, acting as an umbrella term. However, GI itself can have different definitions (e.g., definitions provided by USEPA (2024) or United Nations Information Portal), and since this falls outside the scope of our study, readers are encouraged to refer to Conway *et al.* (2020) and Grabowski *et al.* (2022) for further clarification. It is worth highlighting that the term 'service functions' is used in the following to refer to the services provided by or expected from a GI system (Belmeziti *et al.* 2015), and 'dimensions' refer to distinct categories or aspects, such as environmental, economic, and social considerations, used to analyse and evaluate the sustainability and its impact within the context of GI.

According to Matsler *et al.* (2021), scholars as well as practitioners need to be clear and precise about how they define GI and its purposes in their research. In this regard, the following explanation (Adopted from the *Urban Ecology Concepts* group (Matsler *et al.* 2021)) is provided to the readers based on this study's scope:

By definition, all natural and human-made elements that provide ecological and hydrological functions and processes within, around, and between urban areas, at all spatial scales represent a GI. It can encompass components such as natural heritage features and systems, parklands, stormwater management systems, street trees, urban forests, natural channels, permeable surfaces, and green roofs (Tzoulas *et al.* 2007; Conway *et al.* 2020). The ecosystem multi-functionality of GIs includes protecting against floods, alleviating urban heat islands, reducing greenhouse gas (GHG) emissions, improving air, soil, and water quality, and promoting biodiversity, habitat, and human community well-being. This unique GIs' ability to deliver multiple benefits makes them particularly attractive in urban water management as well as in urban planning. For example, the European Commission has launched a strategy called 'Green Infrastructure – Enhancing Europe's Natural Capital' that intends to mainstream GI in spatial planning and territorial development in order to actively take into account the numerous benefits that humans derive from nature (Hansen & Pauleit 2014). However, GIs planning and evaluation tend to focus on one or a few service functions, mostly related to stormwater management (Finewood *et al.* 2019), failing to optimize (urban) land use. One of the chief reasons could be that only some of the service functions are fairly simple to gauge and assess (Zuniga-Teran *et al.* 2020) or some of the benefits of GIs are often more visible than others (e.g., runoff reduction). However, to build further trust in such new technologies and leverage the appeal of green solutions to different sectors and unlock funding opportunities, a comprehensive performance evaluation of GIs installed is recommended.

Moreover, GI needs ongoing financial support and maintenance as grey infrastructure does in order to fulfil its purpose and offer society the numerous advantages it can (Young & McPherson 2013). To accomplish this objective, it is essential to develop a robust asset management plan. Asset management's main goal is to produce optimal and sustainable financial returns for utility managers and asset owners while simultaneously ensuring that the standards for customer service and safety are upheld (Wenzler 2005). But since GIs show a wide diversity in terms of their type (ranging from bioswales, and green roofs to urban forests), level of multi-functionality, and actors involved in their operation and maintenance (O&M), their asset management poses a more complex challenge compared to traditional grey infrastructures (Langeveld *et al.* 2022). This is why employing a comprehensive and well-defined set of performance indicators (PIs) and a monitoring system based on key service measures has twofold importance in this domain to enhance the effectiveness of asset management strategies.

Performance assessment refers to a broad range of methodologies and techniques that facilitate the evaluation of the efficacy of a process or activity using performance measures. PIs are essential pillars in the assessment process, and they are defined as: 'An item of the information collected at regular intervals to track the performance of a system' (Fitz-Gibbon 1990). According to Ugarelli *et al.* (2021), PIs should be (1) relevant to the strategic objectives, (2) as universal as possible (i.e., not case-dependent), (3) composed in a clear and concise way, and (4) be straightforward and simple to understand. In the current approach to GI performance assessment, only one or a few important PIs are examined (Fu *et al.* 2021). However, a key element of the successful adoption and implementation of GIs as an alternative to grey infrastructure is the comprehensive assessment of the advantages they offer. Otherwise, GI may appear to be less effective than traditional grey infrastructure (Alves *et al.* 2019). It is very rare to find studies indicating how and how much GIs affect intangible benefits such as biodiversity preservation, quality of life and well-being, aesthetics, and recreation (Manso *et al.* 2021). One of the factors of this deficiency can be the lack of a comprehensive performance evaluation system (including well-defined PIs and how to monitor them). Without proper guidelines, asset managers lack standardized procedures to assess strategic, tactical, and operational objectives, and decide intervention alternatives in terms of performance, risk, and cost, over the GI's lifetime (Alegre & Coelho 2012).

In recent years, several national and international guidelines for evaluating the benefits of GIs have been developed to aid in the selection of the most appropriate performance and impact indicators. They can aid in ensuring that evaluations of the performance of GI are uniform and consistent among various projects, localities, and stakeholders. However, there is a strong need for reviewing guidelines to compare considered PIs, relevance and accuracy of suggested PIs, consistency, and understandability of the guidelines, etc. Such a review will enable the improvement of existing guidelines in order to provide a more comprehensive list of PIs, enhance transparency in their evaluation, and adopt a more robust methodological approach. The review also establishes the level of compatibility between guidelines and thus helps the end users (mainly policymakers, utility managers, and municipalities) to determine which guidelines are compatible for benchmarking. A lack of clear and practical answers to these issues could jeopardize the commitment of end users to apply the developed guidelines. This

study aims to critically examine guidelines from Asia, Australia, Europe, North and South America and compare them, in order to evaluate their state of completeness in dealing with various functional aspects of GIs. Furthermore, the quality of existing guidelines is evaluated by comparing them against a set of rigorous quality criteria for guideline development.

2. MATERIAL AND METHODS

The analysis of the guidelines involves evaluating different chapters for two main purposes: first, to assess the comprehensiveness of the guidelines in addressing the PIs, and second, to evaluate their level of development quality. Figure 1 illustrates the stepwise process of evaluating the chapters. These steps are described in detail in the following sections.

2.1. Assessment of the comprehensiveness of guidelines concerning the number of addressed PI

First, publicly available literature and materials were reviewed to extract suggested PIs for assessing different service functions. The search strategy involved using terms such as 'performance indicator', and 'performance assessment', in conjunction with key frequently used terminologies for these natural infrastructures such as 'green infrastructure' and 'nature-based solutions' to identify relevant scientific articles presenting one or more PIs, in databases like Web of Science and Scopus. Search results encompassed a range of sources, including but not limited to Pakzad & Osmond (2016), Gordon *et al.* (2018), Sun *et al.* (2020), and Sánchez *et al.* (2022). In total, 39 PIs were extracted and classified into four broad dimensions: Catchment sustainability, Economic, Environmental, and Social. Service functions in the Catchment sustainability area primarily address stormwater quantity management and water resource protection. It should be emphasized that water management remains the key reason behind GI adoption in urban areas and GIs are increasingly linked to urban water management (Fletcher *et al.* 2015; Liu & Jensen 2018). As such, this group of service functions with their corresponding indicators is presented separately. Economic growth/savings and energy conservation service functions are covered in the Economic dimension. Service functions of GI with respect to the environment can be divided into four categories: air quality, ecological services, water quality, and soil quality. Social service functions of GI include measures related to community liveability and well-being, research and education, and equality and justice. Table 1 presents the complete list of indicators along with their corresponding functional classification, where 'Level 1' and 'Level 2' represent, respectively, the general and specific service

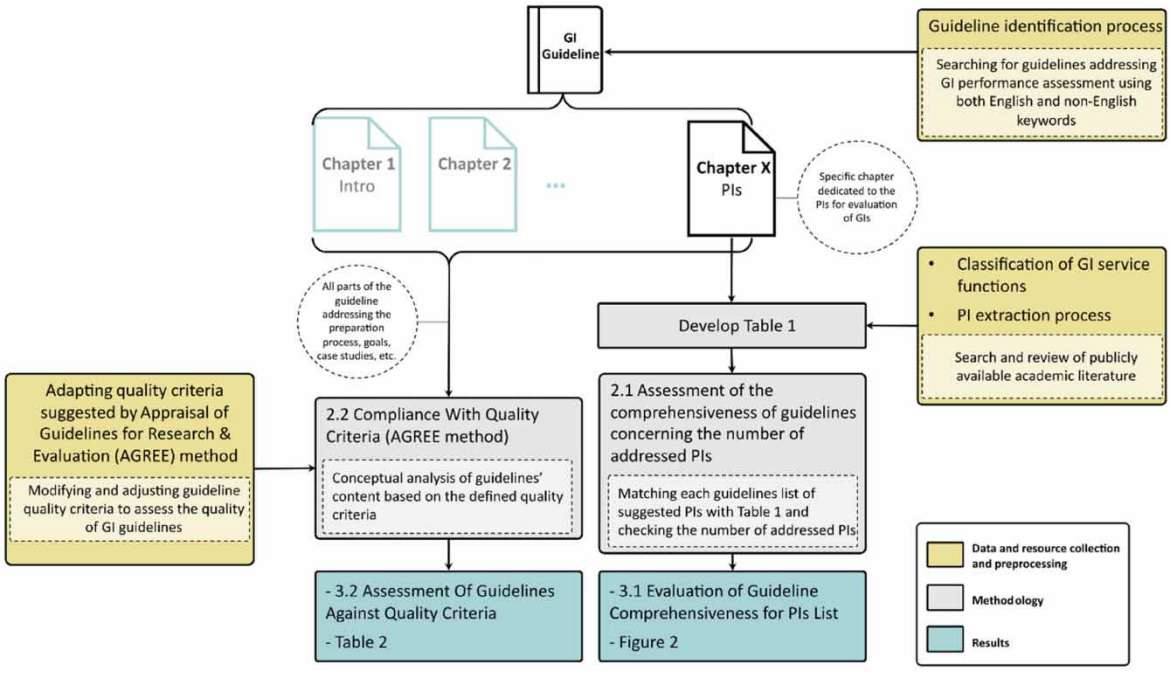


Figure 1 | Proposed stepwise process for evaluating the comprehensiveness of guidelines in addressing performance indicators (PIs) and their level of development quality.

Table 1 | Summary of GIs' impacts dimension, service functions, and suggested PIs in the reviewed literature (hierarchical structure inspired by Pakzad & Osmond (2016))

Dimensions	Service functions (level 1)		Service functions (level 2)	Performance indicator	Symbol	Reference
Catchment sustainability	Managing stormwater runoff		Reduction of flooding risks	(Peak) flow attenuation (%)	CS1-1	Lucas & Sample (2015)
			Interception of runoff	Runoff volume reduction/ retention (m^3)	CS1-2	
			Combined sewer overflows (CSO) control	No. of CSOs per year	CS1-3	
	Protection of water resources		Groundwater recharge	The water level in wells (m)	CS2-1	Healy & Cook (2002)
			Irrigation offset	Decreased water use for landscaping (m^3/year)	CS2-2	Xiao *et al.* (2007)
Economic	Economic growth/saving		Residential property value enhancement	Mean land and/or property value in proximity to GI ($)	Ec1-1	Dell'Anna *et al.* (2022)
			Raise city revenue	No. of tourists	Ec1-2	Yu (2020)
			Boost in employment	No. of created green jobs	Ec1-3	Jarvis *et al.* (2011)
			Reduction of grey infrastructure needs	Grey infrastructure expansion/repair cost reduction ($)	Ec1-4	Xu *et al.* (2019)
	Energy	Saving	Alleviation of energy consumption in (waste)water treatment plant	Decreased volume of (waste) water from the treatment system (m^3/year)	Ec2-1	Sabia *et al.* (2020)
			Reduction of energy consumption needed for cooling and heating	Energy cost reduction ($/year)	Ec2-2	Chang *et al.* (2011)
Environmental	Water quality	Pollution control	Nutrients control	% Of removal of phosphorus and nitrogen	WQ1-1	Fan (2020)
			Reduction/elimination of microbial contamination	% Of removal of Heavy metals, pesticides, etc.	WQ1-2	Fowdar *et al.* (2021)
			Hazardous/toxic material control	% Of the removal of Total suspended solids (TSS)	WQ1-3	Lundgren (2021)
			Sediment transport and/or TSS reduction	Turbidity in streams (NTU)	WQ1-4	Phillips *et al.* (2022)
	Ecological services	Microclimate modification	Pedestrian comfort	Monthly mean value of daily max (min) temperature (°C)	ES1-1	Rosenzweig *et al.* (2006)
			Surface temperature reduction	Land surface temperature (°C)	ES1-2	Maimaitiyiming *et al.* (2014)
			Wind attenuation	Wind comfort level (m/s)	ES1-3	Ricci *et al.* (2022)
		Blue-green space improvement	Greened and/or recreational areas augmentation	Tot. surface area of blue/green spaces (ha)	ES2-1	Kabisch & Haase (2013)

(*Continued.*)

Table 1 | Continued

Dimensions		Service functions (level 1)	Service functions (level 2)	Performance indicator	Symbol	Reference
			Vegetation cover	Percentage of tree canopy cover (%)	ES2-2	Ziter *et al.* (2019)
		Wildlife enhancement	Conservation/ restoration of biodiversity (soil and above-ground)	Wildlife diversity (# of species)	ES3-1	Threlfall *et al.* (2017)
			Pollination	Pollinator species presence (#/ha)	ES3-2	Hooftman *et al.* (2023)
		Urban agriculture	Urban food production	Amount of produced crop (ton/ha/year)	ES4-1	Orsini *et al.* (2014)
			Wood provisioning	Quantity of dead wood per unit area (m^3/ha)	ES4-2	Moor *et al.* (2022)
	Air quality	Atmospheric pollution removal	Control of particulate matter	Air quality index	AQ1-1	Mintz (2009)
			Absorption (reduce the emission of) GHGs	The annual amount of captured CO_2 (kg/ha/year)	AQ1-2	Othus-Gault (2021)
	Soil quality	Improving soil physical properties	Enhancing soil moisture storage	Volumetric water content (cm^3/cm^3)	SQ1-1	Verheijen *et al.* (2010)
			Soil erosion control	Soil loss rate (ton/ha/year)	SQ1-2	Fu *et al.* (2011)
		Improvement of soil chemical properties	Enhancement of soil minerals	Soil organic matter (%)	SQ2-1	Tang *et al.* (2021)
Social	Community liveability		Physiological stress relief	Finger pulse rate (bpm)	S1-1	Lin *et al.* (2019)
			Improved aesthetics of the living environment	Property value ($)	S1-2	Dell'Anna *et al.* (2022)
			Increased human health and well-being	Prevalence of cardiovascular disease (%)	S1-3	Seo *et al.* (2019)
			Development of social interaction/ connectivity	Hours of social activities	S1-4	Peschardt *et al.* (2012)
			Lowering crime rate	No. of violent incidents, nuisances and crimes (per 100,000 population)	S1-5	Lopes & Camanho (2013)
			Public accessibility	Walking distance to the nearest green or blue space	S1-6	Du *et al.* (2020)
			Disaster resilient society	Decrease in loss of life and property during flood events ($)	S1-7	Sohn *et al.* (2021)
			Noise reduction	Noise reduction coefficient	S1-8	Connelly (2011)
	Research and education		Citizen involvement in environmental learning opportunities	No. of environmental education excursions	S2-1	Wolsink (2016)

(*Continued.*)

Table 1 | Continued

Dimensions	Service functions (level 1)	Service functions (level 2)	Performance indicator	Symbol	Reference
	Equality and justice	Increased social inclusiveness of excluded groups (e.g., elderly, persons with disabilities, immigrants, etc.)	Proportion of targeted group (e.g., elderly residents) among the residents (%)	S3-1	Ali *et al.* (2022)

functions. These two columns were designed and categorized by the authors, drawing inspiration from studies such as Ommer *et al.* (2022). Concerning the last column, it's worth noting that only one reference is provided as an example. However, there may be additional references in the literature.

Regarding the guidelines, as illustrated in Figure 1, a similar search strategy to that of PIs was employed: the authors conducted a thorough search of online documents using search engines, employing keywords such as 'nature-based solutions,' 'green infrastructures,' etc., in conjunction with terms like 'performance assessment guideline' and 'performance assessment standard'. Additionally, to ensure the inclusion of non-English documents, the search was repeated using translated keywords in various languages, such as Spanish. It is important to emphasize that by 'performance assessment guideline' the authors sought guidelines that included a list of PIs. As a result, five guidelines from five different regions around the globe were detected: New South Wales (NSW) in Australia, China, Colombia, Europe, and the USA, which encompass various target communities ranging from the state to federal level.

The state of NSW in Australia, as one of the pioneers' states in examining co-benefits of GIs, has published a framework for valuing public space and GI (NSW Department of Planning and Environment 2022). The framework can be used to prepare economic evaluations, including cost–benefit analyses. The other chapters discuss the purposes behind developing the framework, GIs benefits and implementation costs, sensitivity analysis for risks, limitations of the proposed framework, case studies, and non-market valuation of GIs.

In China, the Ministry of Housing and Urban–Rural Development (MOHURD) published the 'Sponge City Performance Evaluation Index System' (Wang *et al.* 2020) and divided the evaluation into four grades (from 'Excellent' to 'Disqualified'). Besides the index system, seven items, including runoff volume control, water quality, and ecological conservation are used to assess project benefits. These criteria were selected to comply with the defined development goals of Sponge City which are, in turn, based on the philosophy of 'source reduction, process control and systematic remediation' (Wang *et al.* 2020). Furthermore, the guideline discusses basic requirements for implementation and methods of assessment for proposed PIs.

Figueroa Arango (2020) specifically addressed the Colombian environment when recommending indicators that should be considered to evaluate and monitor GIs. The guideline is organized into seven steps with the last step dedicated to various types of indicators to monitor GIs and classifies them into seven broad groups. The other steps discuss GI opportunities for urban areas, environmental challenges addressed by GIs, consideration of future urban development scenarios, prioritization of urban areas for GI implementation, multifunctionality and financial strategies for carrying out GI projects.

A handbook on assessing NBS performance and impact was published by the European Commission in 2021 (Dumitru & Wendling 2021). The handbook is designed to apply to GIs implemented across a wide geographic area and at a wide scale. It provides decision-makers with a set of indicators and methodologies to evaluate the impacts of GIs across 12 societal challenges, including climate resilience, water management, biodiversity, etc. Furthermore, the handbook examines GI monitoring strategies, the application of the proposed methodology on case studies, GIs role in risk reduction, and data requirements for GI assessment.

In the USA, AECOM prepared the 'Green infrastructure resource guide' (AECOM 2017) for the United States Agency for International Development (USAID). They categorized the benefits of GIs in 11 classes and for each of them provided relevant indicators to aid in monitoring the performance of a GI solution. The guide is presented in three main sections, discussing GI benefits as a sustainable approach to environmental stressors, moving on to discussing considerations to maximize GI benefits, and implementing GIs in different settings and scales. It should be noted that each of the reviewed sources has used a specific word in its title (such as a

handbook, guideline, or standard). However, upon comparing the definitions of these words in the dictionary (e.g., Oxford English Dictionary 2003), it becomes clear that they all refer to a similar concept: rules or instructions provided by an official organization that tell you how to do something and make judgements. Therefore, in the present text, the term 'guideline' will be used to encompass all of them.

2.2. Compliance with quality criteria

A guideline's potential benefits are limited by its quality. Guidelines of superior quality are founded on both empirical evidence and a widespread consensus of viewpoints, which in turn, expedites their adoption and effective utilization within the intended audience (Grol *et al.* 1998). Regarding methodological development, the validity of the guideline elaboration process ensures the validity of the guidelines. To achieve this level of quality, guidelines must therefore be developed using a structured and coordinated program that follows the principles of evidence-based guideline development (Burgers *et al.* 2003). Also, explicitly reporting the methods and procedures used in their development can further enhance their quality (Fervers *et al.* 2005). However, often, guidelines are not up to basic standards, and the quality can be extremely variable.

In this regard, a framework for assessing the quality of clinical practice guidelines was developed, known as the Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument (AGREE Next Steps Consortium 2017). More precisely, according to the AGREE Collaboration, the quality of guidelines refers to whether the potential biases in guideline development have been adequately addressed and whether the recommendations are both internally valid and externally credible (Xie & Wang 2012). Presently, the tool is accessible in over 10 languages and is being utilized in various nations across the globe (Burgers 2006). AGREE has six domains (i.e., aspects of the guideline that are examined) which are briefly described as follows:

- Domain 1 (D1), Scope and Purpose, focuses on the overall aim of the guideline, the geographical area where it can be applied, the intended audience, and other related criteria.
- Domain 2 (D2), Stakeholder Involvement, considers the degree to which the guideline was developed by relevant stakeholders and represents the views of its intended users.
- Domain 3 (D3), Rigour of Development, pertains to the process used to gather and synthesize indicators, the assessment process, the interpretation of evaluated PIs, and the guideline update procedure.
- Domain 4 (D4), Clarity of Presentation, deals with the language, structure, and format of the guideline.
- Domain 5 (D5), Applicability, concerns the potential barriers and facilitators to implementation, as well as the assessment procedure.
- Domain 6 (D6), Editorial Independence, is focused on ensuring that recommendations are not unduly biased by competing interests.

Each Domain includes several criteria (principles by which guideline quality is judged with respect to each domain) which are modified and adjusted (wherever necessary) in this paper, to assess the quality of GI performance assessment guidelines developed by local, regional, national, or international groups or affiliated governmental organizations. For example, authors have moved 'The target users of the guideline are clearly defined' criterion from the 'Stakeholder involvement' domain to a more relevant domain 'Scope and purpose'. Or, for another example, the criterion called 'The health benefits, side effects, and risks have been considered in formulating the recommendations' is omitted as it is irrelevant in our context. Or for some criteria, simply the clinical terms are replaced by infrastructure-relevant terms (e.g., specifications of GIs instead of patients in 'The patients to whom the guideline is meant to apply are specifically described'). Nevertheless, it should be emphasized that the majority of criteria proposed by AGREE (e.g., clarifying objectives, update procedures, etc.) are not specific to the guideline's topic and should be applicable to any type of guideline. Ultimately, 19 redefined/intact criteria were considered under six general categories. In the following section, detailed descriptions for each criterion are provided.

C1 (D1) – The overall objective(s) of the guideline is (are) specifically described.

The overall objective(s) of the guideline should be described in detail and the expected benefits from the guideline should be specific to GIs performance assessment topic. For example, specific statements would be:

1. Facilitate the adoption of consistent indicators and methods to assess the performance and impact of various types of GIs.
2. Assist urban practitioners in designing effective impact evaluation frameworks for GIs.
3. Establish scientifically sound monitoring and evaluation plans to evaluate the impact of GIs.

C2 (D1) – The main terms and principles in GIs performance and impact evaluation are specifically described.

In this context, it is important to provide explanations for key terms such as performance, impact, monitoring, and evaluation process. Additionally, the necessity of performance and impact evaluation in facilitating decision-making, description of the approach in process and methods, assumptions and evidence used, etc., should be elaborated upon. Furthermore, it would be beneficial to address certain questions such as 'What steps are involved in developing a reliable and suitable impact evaluation framework?'

C3 (D1) – Specifications of GIs to which the guideline is meant to apply are specifically described.

A clear description of the GIs (e.g., type, size, location, etc.) covered by a guideline should be provided. For example, one classification based on the level and magnitude of impacts can be (adopted by Dumitru & Wendling 2021):

Type 1: involves little to no intervention in ecosystems and aims to maintain or enhance the provision of ecosystem services.
Type 2: employs intensive or extensive management strategies to establish sustainable, multifunctional ecosystems and landscapes that enhance the provision of ecosystem services.
Type 3: involves highly intensive ecosystem management practices or the creation of new ecosystems to achieve specific objectives.

C4 (D1) – The target users of the guideline are clearly defined.

The target users should be clearly defined in the guideline, so the reader can immediately determine if the guideline is relevant to them. For example, the target users of a guideline on GIs impact assessment may include field practitioners, planners, and decision-makers who implement these green infrastructures.

C5 (D2) – The guideline development core group includes individuals from all the relevant professional groups.

This criterion refers to professionals who participated in the development process, which could encompass members of the steering group, the research team responsible for selecting the indicators and evaluating evidence, and those involved in creating the final recommendations. Excluded from this item are individuals who externally reviewed the guideline (refer to C10) as well as target population representation (refer to C6). It is highly recommended to provide details about the composition, discipline, and relevant expertise of the guideline development group.

C6 (D2) – The views and preferences of the end-users have been sought.

It is highly beneficial that stakeholder organizations contribute to and comment on the PI assessment guideline at various stages. For instance, after defining exactly what the guideline will and will not cover, guideline developers should arrange a workshop for all stakeholder organizations and individuals (e.g., municipalities, water utilities, etc.) to present the under-development standard. The merits of the guideline could be discussed by the attendees followed by posting the draft online for consultation (for a period of e.g., 1 month). Moreover, after publication, stakeholders can act as the implementation support and encourage their networks to use the guideline at both national and local levels.

C7 (D3) – The proposed indicators, assessment frameworks, etc. have been selected and developed based on evidence.

To ensure a thorough search for evidence, it is necessary to provide specific details of the method and strategy employed, including the search terms, sources consulted, and the dates of literature searched. These sources may include electronic databases (such as high-impact journal articles), databases of systematic reviews, conference proceedings, and other similar guidelines. The search strategy should aim to be as comprehensive as possible while avoiding potential biases, and the execution of the search should be well-documented and sufficiently detailed.

C8 (D3) – The strengths and limitations of the proposed PIs/assessment framework are clearly described.

Statements highlighting the strengths and limitations of the assessment framework and the applicability of the proposed PIs should be provided. Although the proposed methodology/PI set should be objective, standardized, and well-adapted to different contexts, some questions like the following should be addressed:

1. Can they be used for both post-ante and ex-ante evaluation of the infrastructure?
2. Are there any indicators that may not be applicable to specific environments?
3. How does the proposed framework deal with data uncertainty?

C9 (D3) – The way to interpret calculated/estimated PIs is presented in a clear and comprehensive manner.

To assess performance, PI values must be compared to reference values, which can be accomplished by creating performance functions for each PI. These functions establish a correlation between PI values and a classification scale, allowing for the identification of good, satisfactory, and poor performance. These reference values can be defined based on legislation, literature references, etc.

C10 (D3) – The guideline has been externally reviewed by experts prior to its publication.

Before publication, it is important to subject a guideline to external review. To ensure impartiality, the reviewers should not have been part of the guideline development group. The review panel should consist of experts in urban infrastructure management as well as methodological experts. End users such as policy makers and urban planners may also be invited to provide input. The methodology employed in the external review should be clearly outlined, including a list of the reviewers and their affiliations.

C11 (D3) – A procedure for updating the guideline is provided.

Guidelines should be dynamic, requiring an annual review of their scope, suggested indicators, assessment procedures, and so on, which involves conducting an updated literature search to identify new evidence and accounting for new factors and obstacles encountered in practice. This is very important, especially in the field of GI, where specialized knowledge is rapidly evolving.

C12 (D3) – Strategies for disseminating guidelines to end users are established.

According to Boulet *et al.* (2006), an effective approach necessitates the development of a dissemination plan that is ideally generated concurrently with the production of the guideline; the plan should be made during the project and not at the end of the project. The target audience must be identified from the project's outset to not only decide the project's scope, objectives, format, style, and wording for the recommendations but also for choosing the best instruments for dissemination (Eccles *et al.* 2012). The use of both conventional and cutting-edge dissemination technologies should enhance dissemination efforts.

C13 (D4) – The proposed PIs to assess the potential benefits and co-benefits of GIs are classified properly and are specific and unambiguous.

The guideline should present PIs in well-defined categories (based on the benefits and co-benefits of GIs) and provide the reader with a concrete and precise description of each indicator. Moreover, units of measurement for each PI must be clearly stated.

C14 (D4) – The possible alternative PI(s) for each of the different performances are explicitly presented.

A guideline aimed at evaluating the performance of GIs should incorporate a range of options for assessing specific (co)benefits, to be more resilient. Moreover, the end-user can choose the right indicator depending on the available facilities and the project goal. These alternatives should be clearly outlined within the guidelines. For instance, in evaluating the benefits of GIs for reducing runoff quantity, the guideline might consider a variety of indicators, such as the Run-Off Score, Flood Excess Volume, Height of Flood Peak, and Peak Flow Variation.

C15 (D5) – The guideline describes facilitators and barriers to its application.

The successful implementation of guidelines hinges on addressing various barriers (such as personal, performance evaluation process, etc.) and implementing effective strategies to overcome them. Additionally, factors that can promote guideline implementation or adherence should be addressed. These facilitators include the dissemination and advertisement of guideline materials, educating and training individuals on the guidelines, regulatory and financial incentives, and institutional support.

C16 (D5) – Pilot testing of guidelines before widespread implementation.

A pilot test offers the chance to test out a new procedure on a limited scale while getting feedback. As a result, any process flaws can be fixed before being applied over the entire facility. Although this method will postpone publication, it might guarantee the project's success. One should think about the following issues before starting the pilot:

1. Which locations are preferable to test the procedure in?
2. Are the locations selected for the pilot already involved and supportive of it?
3. What system will be in place during the trial to allow front-line employees to provide and receive feedback?
4. What are the process measures being used to gauge compliance during the pilot, such as quantity and adherence to the process?

C17 (D5) – The guideline provides measurement advice and/or tools on how the PIs should be calculated and assessed.

A guideline must be implemented with supporting documents on the evaluation process to be effective. Examples of these can include the scale of measurement, frequency of monitoring, instructional resources, the findings of a pilot study, or computer assistance.

C18 (D6) – The guideline was developed independently and without any bias or undue influence from the funding body. Many guidelines are created with outside financing (e.g., government, professional associations, etc.). Support can take the form of a monetary donation towards the entire development or just certain aspects of it (e.g., the printing of the guidelines). There should be a clear declaration that the views or interests of the funding entity have not influenced the final manuscript and recommendations.

C19 (D6) – Competing interests of guideline development group members have been recorded and addressed.

Members of the development group can have conflicting interests. This might be the case, for instance, if a member of the development group also receives funding from another organization for research on the subject addressed by the guideline. There should be a clear declaration that each group member has disclosed any competing interests.

In the next step, we analysed the lists of performance assessment indicators and also reviewed the content from other chapters of different guidelines for the GIs performance assessment procedure to determine: (1) their proposed indicators to assess the performance of a GI project and their comprehensiveness across a range of social, environmental, and economic contexts; and (2) their compliance with the described quality criteria.

3. RESULTS AND DISCUSSION

3.1. Evaluation of guideline comprehensiveness for PIs list

Based on the comparison of the performance assessment indicators, the EU guideline is the most comprehensive one considering the number of addressed PIs (see Figure 2). Only a few indicators like 'Grey infrastructure expansion/repair cost reduction' are missing in the current version. However, the term 'indicator' is not well-defined

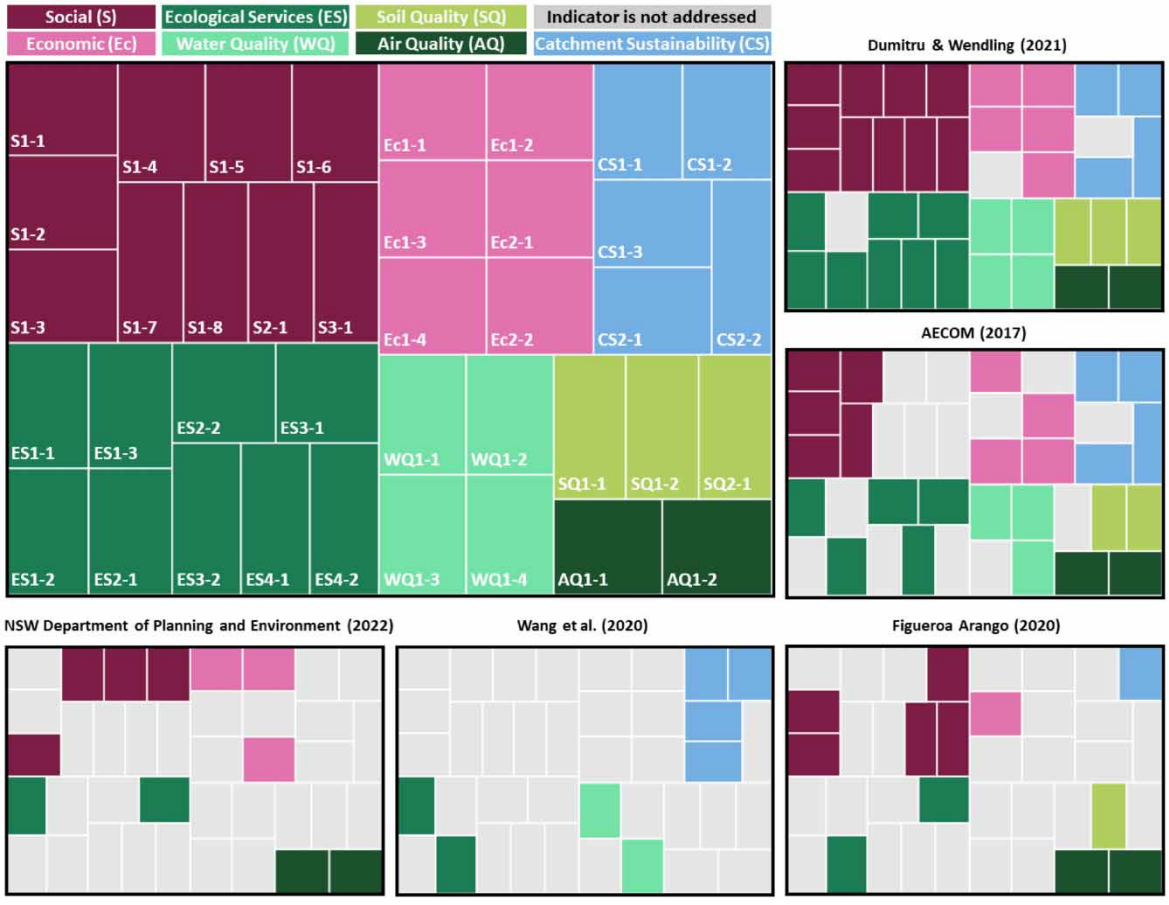


Figure 2 | Performance indicators addressed in each of the studied guidelines.

and sometimes mixed up with 'service functions of GIs' and resulting in a long indicator list. As an example, 'Citizen involvement in environmental education activities' is considered a PI (while it is a service function). As suggested by Silva *et al.* (2023) a high number of indicators may bring serious challenges to the assessment procedure. Nevertheless, as a strength point, this guideline distinguishes between the list of core and additional indicators. Another important issue is that some indicators are not explicitly defined, making their assessment in the current state difficult and subject to variation for each application (e.g., 'Encouraging a healthy lifestyle'). Following the European document, the USAID guideline has the second-most comprehensive PI list by addressing nearly two-thirds of the indicators. However, in this document, PIs related to social dimension are somewhat neglected. This negligence can result in a failure to clarify the positive effects of GIs in this area. In addition, ignoring the social dimension can result in unfair outcomes because different social groups may not equally benefit from GI investments. In this guideline also some of the indicators are defined in an overly general way: e.g., 'Air quality' or 'Air temperature variations'. It might be challenging to figure out if a project is succeeding when PIs are too general since they are hard to monitor and interpret. As a strength, in this guideline benefits gained from GIs are colour-coded to identify the highest-impact design options. Also, all the PIs are combined with up/down arrows to show the direction of the index's desirability.

In the NSW guideline, everything is translated into the monetary scale (e.g., value in \$ per m^2 of tree canopy); which can facilitate the decision-making process, especially in cost–benefit analysis. However, the cost–benefit approach also requires a fair financial valuation of non-monetary benefits such as well-being or biodiversity preservation. The most serious shortcoming of this guide is not addressing some of the significant benefits of these infrastructures, such as 'Catchment Sustainability'. One of the main drawbacks is that it may lead to insufficient knowledge of the advantages of GI for water quality and quantity management. The full value of GI cannot be understated if any of these key service functions are not considered throughout the evaluation process. In addition, disregarding these factors may have unforeseen consequences that could compromise the project's sustainability, such as decreased water quality, increased runoff, and increased flooding. Neglecting key roles of GIs in urban water management might also limit the possibilities for innovation and advancement in their design and application. In contrast, the performance assessment standard for the Chinese sponge cities has effectively addressed 'Catchment sustainability' and 'Water quality' related indicators by incorporating a greater number of these indicators. However, it avoids evaluating the microbial and toxic material concentration. On the other hand, this document does not inscribe any of the 'Social', 'Air quality' and most of the 'Ecological services' related indicators. In other words, this guideline is more centred around the water management benefits of GIs. As was already indicated, one of the main consequences of neglecting other crucial GI service functions is that the benefits of these infrastructures as a whole may not be fully understood. Beyond just managing urban runoff, GI can also have positive social and economic effects as well as improve the natural environment. The full benefit of the GI project might not be realized if these other service functions are not considered during the performance evaluation process. Besides, decision-makers risk missing opportunities to create more effective and efficient solutions that tackle several problems at once if they just take into account a small number of service functions. It should be highlighted that GIs can be constructed to meet the particular needs of the society, but having access to thorough performance assessment guidelines is crucial to do comprehensive performance assessments whenever necessary.

Colombia's guideline makes an effort to reflect various benefits of GIs while proposing the PIs; however, there is still room for improvement in defining the PIs. As an example, for the 'Soil erosion control' service function, instead of 'Area conserved to prevent erosion processes' which might seem slightly unclear to the decision maker, 'Soil loss rate' can be adopted as a PI or instead of 'Planting trees to reduce the effect of the urban heat island', 'Land surface temperature' can be used as a PI. In other terms, PIs should be result-oriented instead of mean-oriented. Moreover, some of the suggested PIs are not related to monitoring the performance of a GI itself but are rather suitable to evaluate the performance of the city planners/managers; like 'Number of GIs that include participation of different stakeholders in the design'. Besides, this guideline emphasizes assessing the social and (partially) environmental performances of GIs and neglects their other important beneficial categories like catchment sustainability and water quality. In general, using an incomplete or imperfect performance indicator list can lead to incorrect conclusions about infrastructure projects' quality and effectiveness. This results in an assessment that favours certain aspects of infrastructure performance over others. The ability to hold developers and operators accountable for their performance can also be limited as a result. Accordingly, infrastructure development processes may be complicated and lacking in transparency and accountability.

The authors believe that variations in the perceptions and expectations of regions towards the development of GIs might be responsible for the differences in the PIs addressed in each of the guidelines. For instance, sponge cities in China, as indicated by their name, place a strong emphasis on the advantages associated with both the quantity and quality of water. However, an interesting observation about the terminology used for these infrastructures can be found in the four guidelines. Both USAID and NSW utilize 'GI' in their title, while EU and the Colombian guidelines use 'NBS'. However, each pair has a distinct perspective on the respective terms: EU and USAID encompass a variety of PIs related to diverse service functions, whereas regions like NSW or Colombia prioritize traditional benefits of GIs over hydrological benefits. This finding demonstrates that even using the same terminology to refer to these natural infrastructures may result in different translations across various regions worldwide. Additionally, the level at which these standards were compiled and published can also be a significant factor contributing to these differences. In other words, certain standards are established at the federal level (e.g., EU or USA), while others are developed at the national level (e.g., Colombia or China), and some may be specific to the state level (e.g., NSW). Here, also a question may arise: would it be more beneficial to develop separate performance assessment guidelines, specialized for each specific GI dimension, offering a more targeted assessment? Or alternatively, should performance assessment guidelines be developed in a unified, comprehensive way to make their implementation more practical by providing consistent terminology and assessment methodologies? Recognizing the diverse decision levels in the field, the authors believe there is a need for distinct guidelines for strategic and operational levels of GI management. At the strategic level, a more comprehensive set of indicators could offer decision-makers a more complete overview of GI service functions and the resources that might be needed to assess each of those performances, without delving too deeply into the performance assessment procedure itself. However, at the operational level, more focused and pragmatic performance assessment guidelines are necessary, focusing on further details of GI typologies and components that provide such services, and providing details on the measurement and assessment procedures of PIs addressing that specific service function.

Regarding the PI names, it should be noted that in some cases, different guidelines translated similar indicators differently: for instance, the 'Prevalence of cardiovascular disease' indicator is suggested by the EU standard for evaluating 'Human health and well-being' while the NSW standard uses 'Physical activity and morbidity rate', USA standard utilizes 'No. of doctor visits' and in Colombia 'Number of people using the GI to do sports' is adopted. In this case, as well as similar cases, the PI was considered addressed. Moreover, for some of the GI's service functions, there is more than one indicator suggested by the literature (e.g., for 'Managing stormwater runoff': 'Surface runoff in relation to precipitation quantity', 'Peak flow variation', 'Flood peak reduction', 'Height of flood peak', 'Runoff score', etc.). Considering the available budget, the purpose of the project, the availability of data, and so on, one or more of them can be adopted by the practitioners. Furthermore, it is neither possible nor logical to evaluate some of the extracted indicators at the individual infrastructure level and for all types of GIs. For instance, bioswales may not have any special effect on tourism or the assessment of GHG reduction makes sense when several GIs function together. Generally, it is important to select the right set of indicators, depending on the decision context in which the indicators will be used.

3.2. Assessment of guidelines against quality criteria

For the evaluation of the guideline quality, the guidelines documents were carefully read and evaluated against each defined criterion. Prior to the appraisal, the authors also attempted to collect all information about the guideline development process. Since, in addition to the guideline main manuscript, this information may also be included in a separate technical report, ethical manual, website, or policy statement for guideline developers. The content of guidelines was examined against quality criteria using qualitative content analysis in this study. For further information on this method, readers are encouraged to refer to relevant literature such as Bengtsson (2016). Table 2 provides a summary of how well the five studied guidelines conform to the selected criteria. Although the overall goals under Scope and Purpose (Domain 1) have been satisfactorily met by all standards, the American and Colombian guidelines did not outline the key concepts in assessing different performances and impacts of GIs. Failure to clearly describe the processes, methods and assumptions used in the guidelines may result in inconsistent interpretation of the guidelines. This can lead to confusion and inconsistency in applying the guidelines, resulting in sub-optimal results. The characteristics of GI that the guideline should address were not also covered in the Chinese and Colombian standards. If the characteristics of the aimed infrastructures are not properly addressed, the guidelines may be viewed as being irrelevant or inappropriate for the targeted GI,

Table 2 | Compliance with quality criteria of five guideline development programs

Domain	Regional guidelines / Criteria	Dumitru & Wendling (2021)	AECOM (2017)	Figueroa Arango (2020)	Wang et al. (2020)	NSW Dep. Of Plan. & Env. (2022)
1. Scope and Purpose	1. The overall objective(s) of the guideline is (are) specifically described.	+	+	+	+	+
	2. The main terms and principles in GIs performance and impact evaluation are specifically described	+	-	-	+	+
	3. Specifications of GIs (e.g., type, size, scale, etc.) to which the guideline is meant to apply are specifically described.	+	+	-	-	+
	4. The target users of the guideline are clearly defined.	+	+	+	-	+
2. Stakeholder Involvement	5. The guideline development group includes individuals from all the relevant professional groups	+	-	~	+	-
	6. The views and preferences of the end-users have been sought	-	-	-	-	+
3. Rigour of Development	7. The proposed indicators, assessment frameworks, etc. have been selected and developed based on evidence	+	-	-	+	+
	8. The strengths and limitations of the proposed PIs/assessment framework are clearly described	+	-	-	-	+
	9. The way to interpret evaluated PIs is presented in a clear and comprehensive manner	-	-	-	-	-
	10. The guideline has been externally reviewed by experts prior to its publication	-	-	-	+	+
	11. A procedure for updating the guideline is provided	~	-	~	-	-
	12. Strategies for disseminating guidelines to end users	-	-	-	-	-
4. Clarity of Presentation	13. The proposed PIs to assess the potential benefits and co-benefits of NBS are classified properly and are specific and unambiguous	+	+	+	+	+
	14. The possible alternative PI(s) for each of the different performances are explicitly presented	+	+	+	+	+
5. Applicability	15. The guideline describes facilitators and barriers to its application	-	-	-	-	-
	16. Pilot testing of guidelines before widespread implementation	+	-	-	-	+
	17. The guideline provides measurement advice and/or tools on how the PIs should be calculated and assessed.	+	-	-	+	+
6. Editorial Independence	18. The guideline was developed independently and without any bias or undue influence from the funding body	-	-	-	-	-
	19. Competing interests of guideline development group members have been recorded and addressed.	-	-	-	-	-

Each criterion is marked with one of the following symbols/colours: (+/green). The guideline (almost) completely addressed the criterion; (−/orange). The guideline poorly/not addressed the criterion; and (~/yellow) The guideline partially addressed the criteria.

hindering consistent application across various infrastructure projects. This can cause stakeholders to become confused and uncertain. Furthermore, it could lead to instructions that are overly general and fail to offer useful advice for under-study infrastructure projects. In general, all Domain 1 criteria were fully covered by European and NSW guidelines.

Most guidelines performed poorly in the second domain, 'Stakeholder Participation', in contrast to the first domain. Only the NSW standard made considerations for the opinions and preferences of end users throughout development. When guidelines are developed without considering end users' preferences, errors and omissions may occur. The end user can provide critical feedback and identify potential issues that the guideline developer

team may have missed, which can lead to better quality guidelines and better results. Furthermore, only the European and Chinese standards attempted to include as many members as possible of all pertinent professional groups in the development of the standard. A guideline that ignores the perspectives and priorities of different professional groups may be inadequately representative of all stakeholders' needs and concerns. Without contributions from a wide range of specialists, the guideline could be biased in favour of particular viewpoints or methodologies. This criterion might be the reason for the incomplete list of PIs for some of the guidelines examined in this study.

The guidelines have often been deficient in the third domain, 'Rigor of Development,' which has the highest number of criteria. More specifically, only the European and Chinese standards, which were supported by data and examples, chose the indicators and their evaluation system based on evidence. Nevertheless, the depth of the literature evaluation differs across the guidelines, with the European guideline conducting systematic reviews of literature and presenting the findings of pertinent studies in evidence tables. Furthermore, the recommended performance evaluation methods' advantages and disadvantages were only noted in the NSW and European standards. If the advantages and disadvantages of the recommended assessment methods are not addressed properly, choosing the best evaluation method for a certain case may become confusing and erroneous for practitioners. A biased judgment to choose an evaluation method could also result from an incomplete discussion of the benefits and drawbacks of the suggested evaluation methods. If just the benefits of a certain approach are emphasized, for instance, it could lead end users to choose that approach even though it is not the best option. Furthermore, only the NSW and Chinese standards were examined by outside reviewers. But nonetheless, it's crucial to disclose the reviewers and the organizations they belong to. In this respect, the Chinese standard is the only one that has done well in this regard by meeting this criterion. Worthy of note, it is crucial to provide the names and organizations of the guidelines' reviewers because this can give the guidelines more credibility. The recommendations of a guideline can be more trusted if one is aware that it has been examined by specialists in the field. Also, guideline reviewers can be held responsible for the quality and correctness of the guideline.

Except for the European and Colombian standards, no guideline mentions the inclusion of an updating process. Nevertheless, in the European standard, there is no mention of the mechanism defined for this purpose. Also, the Columbia standard apparently does not have a continuous update (they planned to test the guideline in some municipalities and update the guideline based on the feedback). If the guidelines are not periodically updated, the assessment methodologies may become outdated and ineffective after a while. Neglecting routine reviews of the performance assessment guidelines and adding more verified cases might also decrease end users' trust in the suitability of the guidelines' recommendations for their requirements. In addition, it appears that none of the guidelines have any strategy for dissemination and encouraging end users to use them in the process of GI monitoring. Without a method for spreading the word about the guideline, potential users might not be aware that it exists. This may lead to a lack of adoption and sparse application of the recommendation, which may reduce its efficacy. Strategies for spreading news can give stakeholders the chance to comment on the guideline and help it to be matured. These chances can be lost if there is no dissemination strategy. A very important criterion (perhaps the most important technical criterion) that has been ignored in this domain by all the guidelines is how to interpret the assessed performance indices. It is not proposed in any of the guidelines how to categorize different performance levels of the indicators. Defects in this section can seriously threaten the effective use of the guideline. This means, for example, decision-makers might not be able to make sound decisions if they are not given instructions on how to interpret the assessed performance indices. Likewise, the results can be misinterpreted. As a result, the level of functioning of the infrastructure may be inferred in the wrong way. It should be mentioned that Silva *et al.* (2023) began to address this issue by gathering/suggesting the appropriate Reference Values for PIs.

In contrast to the previous section, all the guidelines in the fourth area, labelled 'Clarity of Presentation', worked very well and systematically categorized the indicators based on the dimensions and various service functions of GI. Regarding the provision of possible alternatives for PIs, even if the guideline had provided alternative options for some indicators (but not all), it was considered as satisfactorily addressed in this research.

None of the standards have addressed the enablers and barriers that end users can experience while applying in relation to applicability (Domain 5). This omission can make it difficult to apply the guidelines effectively. More precisely, without an awareness of the enhancers and impediments to applying performance assessment methodologies and measuring PIs, resources may be wasted on evaluation approaches that do not effectively satisfy the needs of end users. Before the final version was prepared, some case studies were used to test the European and

NSW guidelines. The piloting of the guidelines can assist in identifying usability problems and potential obstacles to its efficient use. This feedback can be used to improve the guideline's usage and effectiveness. Also, testing the guideline in a pilot program can reveal whether it is scalable and can be successfully applied in a variety of situations and settings. All standards, except for the American and Colombian guidelines, have also offered some measurement advice for end users. It is important to keep in mind that measurement recommendations in guideline manuscripts can encourage standardization in the data collection and reporting process, maintaining consistency across GI projects and enabling comparisons between them. Also, end users can lower performance measurement errors and raise the quality of the data. Additionally, it could aid in making research findings more replicable.

The guidelines all perform poorly in the last domain (Editorial Independence). None of them have addressed the assurance that the viewpoints of financial bodies have not influenced the development and composition of the text. Guidelines and recommendations that end users believe are influenced by financial interests (regulatory capture) may not be taken seriously. This may result in a lack of confidence in the recommendations made by the guidelines and a reluctance to use them when analysing the performance of GIs. It is worth noting that the verification of such cases can be challenging, and only the explicitly provided statement can be considered as evidence in this context. Moreover, guideline developer groups do not publish information regarding their conflicts of interest. The guidelines are intended to be objective and are based on the best available data. If conflicts of interest are not disclosed, the end users may lose faith in the process of developing guidelines.

In general, the European and the NSW standards have considered 12 of the 19 established criteria to evaluate the quality of guidelines, which exceeds the American and Colombian standards, which only cover 5 and 6 criteria, as well as the Chinese standard (which covers 8 criteria). To address why certain criteria or domains are not considered through the development of these guidelines, the authors propose a potential reason: the absence of a well-established and widely recognized 'guideline development process' in the field. As demonstrated in this study, frameworks from other disciplines could be utilized for this purpose, of which GI guideline developers may not be aware. It should be noted that the authors identified instances where certain guidelines only briefly touched on specific criteria, claiming they had been taken into account during the manual development process or would be in subsequent steps. These cases were deemed to have been partially addressed in the study (like 'A procedure for updating the guideline is provided' criterion in the EU guideline). However, it is highly recommended that the guideline development groups explicitly and comprehensively outline all these criteria in the main document (or official appendixes) to increase the guideline quality as well as the trust of the final users.

4. CONCLUSION

The primary focus of this study was to conduct an in-depth assessment and comparison of various guideline development programs in the field of GI performance assessment. The main objective was to assist guideline developers working in the domain of GIs to enhance and systematize the guideline development process. In this regard, information on guideline programs was obtained by examining the handbooks and websites of five organizations from Australia, China, Colombia, Europe, and the USA, as the regions that led the way in GI development initiatives. The guidelines were examined and analysed in two phases.

Comparing the guidelines based on how thorough the suggested PIs were was the first step. The results show that the European Union's guideline is the most thorough, followed by the American guideline. While other guidelines concentrated on assessing GIs' performance in a fewer variety of contexts. For instance, China's guideline focuses on water quality and quantity issues. Nevertheless, as the advantages of GI are numerous and contain social and economic aspects as well, it is highly recommended that performance assessment guidelines also reflect the evaluation of these aspects. Since, if only PIs from certain domains are presented in a guideline, it might encourage project developers and stakeholders to prioritize those ecosystem services over others. This could result in a skewed or incomplete understanding of the subject, as it neglects ecosystem services that lack precise evaluation methods and require dedicated funding for research and development, as suggested by Vaissière *et al.* (2013). By considering all these dimensions in the performance evaluation of GI, end users can make more informed and effective decisions regarding its design, operation, and maintenance. A more thorough assessment of its effectiveness can also aid in raising public awareness and support for the use of GI as a long-term solution to urban problems.

In the second part, a comprehensive set of quality standards comprised of 19 criteria organized into six broad categories was used to assess the guidelines' level of quality. Here, the European and NSW standards

encompassed more modes of maturity than the other guidelines by addressing 12 of the 19 quality criteria. Providing high-quality performance assessment standards for end-users is essential for some important reasons. A high-quality standard guarantees the reliability and consistency of the performance evaluation process. Additionally, giving end users high-quality guidelines can aid in increasing stakeholder trust in the results of the GIs' performance evaluation, including politicians, funders, and the general public.

The findings also indicate that even though a lot of work has gone into developing and compiling these standards, there is still potential for improvement. The following points are recommended to improve the effectiveness of current/future guidelines intended for performance assessment of GIs:

- Considering the inclusion of a range of PIs addressing diverse dimensions within the GI performance assessment guidelines. By doing so, we can better capture the diverse benefits that GIs offer in various contexts to support the successful adoption and implementation of GIs as an alternative to grey infrastructure.
- Development and inclusion of methods to interpret evaluated PIs, since PI as a single number doesn't say much. As it has been pointed out, flaws in this section might substantially jeopardize the guideline's ability to be used effectively; since the outcomes of the performance assessment process can be interpreted incorrectly and it may prevent decision-makers from making well-informed decisions..
- Adhering to the principles of guideline development, such as those recommended by AGREE, and providing end users with enough information about each quality criterion are essential steps. Guidelines that meet quality criteria are perceived as credible and trustworthy by end users, enhancing their acceptance and adoption.

All in all, the results and suggestions from this study can be used as a foundation for increasing the comprehensiveness and quality of the GI performance evaluation guidelines. The forthcoming practical frontier in this field lies in *Capacity Building*. With performance assessment guidelines continually evolving, there's a need for tailored training and capacity-building initiatives. These programs should aim to equip practitioners, decision-makers, and stakeholders with the necessary knowledge and skills to adeptly apply performance assessment guidelines within their specific contexts.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

AECOM 2017 *Green Infrastructure Resource Guide*. United States Agency for International Development, Washington, DC. https://www.usaid.gov/pa00tdbd

AGREE Next Steps Consortium 2017 The AGREE II Instrument [Electronic version]. Available from: http://www.agreetrust.org (accessed March 2023)

Alegre, H. & Coelho, S. T. 2012 Infrastructure asset management of urban water systems. In: *Water Supply System Analysis–Selected Topics* (Ostfeld, A., ed.). InTech Open, Rijeka, pp. 49–74. https://doi.org/10.5772/52377.

Ali, M. J., Rahaman, M. & Hossain, S. I. 2022 Urban green spaces for elderly human health: A planning model for healthy city living. *Land Use Policy* **114**, 105970. https://doi.org/10.1016/j.landusepol.2021.105970.

Alves, A., Gersonius, B., Kapelan, Z., Vojinovic, Z. & Sanchez, A. 2019 Assessing the co-benefits of green-blue-grey infrastructure for sustainable urban flood risk management. *Journal of Environmental Management* **239**, 244–254. https://doi.org/10.1016/j.jenvman.2019.03.036.

Belmeziti, A., Cherqui, F., Tourne, A., Granger, D., Werey, C., Le Gauffre, P. & Chocat, B. 2015 Transitioning to sustainable urban water management systems: How to define expected service functions? *Civil Engineering and Environmental Systems* **32** (4), 316–334. https://doi.org/10.1080/10286608.2015.1047355.

Bengtsson, M. 2016 How to plan and perform a qualitative study using content analysis. *NursingPlus Open* **2**, 8–14. https://doi.org/10.1016/j.npls.2016.01.001.

Boulet, L. P., Becker, A., Bowie, D., Hernandez, P., McIvor, A., Rouleau, M., Bourbeau, J., Graham, I. D., Logan, J., Légaré, F. & Ward, T. F. 2006 Implementing practice guidelines: A workshop on guidelines dissemination and implementation with a focus on asthma and COPD. *Canadian Respiratory Journal* **13** (Suppl A), 5A–47A. https://doi.org/10.1155/2006/810978.

Burgers, J. S. 2006 Guideline quality and guideline content: Are they related? *Clinical Chemistry* **52** (1), 3–4. https://doi.org/10.1373/clinchem.2005.059345.

Burgers, J. S., Cluzeau, F. A., Hanna, S. E., Hunt, C. & Grol, R. 2003 Characteristics of high-quality guidelines: Evaluation of 86 clinical guidelines developed in ten European countries and Canada. *International Journal of Technology Assessment in Health Care* **19** (1), 148–157. https://doi.org/10.1017/s026646230300014x.

Chang, N. B., Rivera, B. J. & Wanielista, M. P. 2011 Optimal design for water conservation and energy savings using green roofs in a green building under mixed uncertainties. *Journal of Cleaner Production* **19** (11), 1180–1188. https://doi.org/10.1016/j.jclepro.2011.02.008.

Connelly, M. R. 2011 *Acoustical Characteristics of Vegetated Roofs – Contributions to the Ecological Performance of Buildings and the Urban Soundscape*. Doctoral dissertation, University of British Columbia. https://doi.org/10.14288/1.0072314.

Conway, T. M., Khan, A. & Esak, N. 2020 An analysis of green infrastructure in municipal policy: Divergent meaning and terminology in the Greater Toronto Area. *Land Use Policy* **99**, 104864. https://doi.org/10.1016/j.landusepol.2020.104864.

Dell'Anna, F., Bravi, M. & Bottero, M. 2022 Urban green infrastructures: How much did they affect property prices in Singapore? *Urban Forestry & Urban Greening* **68**, 127475. https://doi.org/10.1016/j.ufug.2022.127475.

Du, X., Zhang, X., Wang, H., Zhi, X. & Huang, J. 2020 Assessing green space potential accessibility through urban artificial building data in Nanjing, China. *Sustainability* **12** (23), 9935. https://doi.org/10.3390/su12239935.

Dumitru, A. & Wendling, L. (Eds.). 2021 *Evaluating the Impact of Nature-Based Solutions: A Handbook for Practitioners*. European Commission EC, Luxembourg. https://doi.org/10.2777/244577.

Eccles, M. P., Grimshaw, J. M., Shekelle, P., Schünemann, H. J. & Woolf, S. 2012 Developing clinical practice guidelines: Target audiences, identifying topics for guidelines, guideline group composition and functioning and conflicts of interest. *Implementation Science* **7** (1), 1–8. https://doi.org/10.1186/1748-5908-7-60.

European Commission, Directorate-General for Research and Innovation. 2015 *Towards an EU Research and Innovation Policy Agenda for Nature-Based Solutions & Re-naturing Cities: Final Report of the Horizon 2020 Expert Group on 'Nature-Based Solutions and Re-naturing Cities'*. https://data.europa.eu/doi/10.2777/479582.

Fan, H. 2020 *Impact of Green Infrastructure on Nutrients Reduction, the Case of Downtown Durham, North Carolina. Master's Project*. Duke University, Durham, NC. Retrieved from https://hdl.handle.net/10161/20560

Fervers, B., Burgers, J. S., Haugh, M. C., Brouwers, M., Browman, G., Cluzeau, F. & Philip, T. 2005 Predictors of high quality clinical practice guidelines: Examples in oncology. *International Journal for Quality in Health Care* **17** (2), 123–132. https://doi.org/10.1093/intqhc/mzi011.

Figueroa Arango, C. 2020 *Guía Para la Integración Delas Soluciones Basadas en la Naturaleza en Laplanificación Urbana*. Primera aproximación para Colombia, Instituto Humboldt, Colombia. https://www.ecologic.eu/17623

Finewood, M. H., Matsler, A. M. & Zivkovich, J. 2019 Green infrastructure and the hidden politics of urban stormwater governance in a postindustrial city. *Annals of the American Association of Geographers* **109** (3), 909–925. https://doi.org/10.1080/24694452.2018.1507813.

Fitz-Gibbon, C. T. 1990 *Performance Indicators (Vol. 2)*. Multilingual Matters, Clevedon, UK.

Fletcher, T. D., Shuster, W., Hunt, W. F., Ashley, R., Butler, D., Arthur, S., Trowsdale, S., Barraud, S., Semadeni-Davies, A., Bertrand-Krajewski, J. L., Mikkelsen, P. S., Rivard, G., Uhl, M., Dagenais, D. & Viklander, M. 2015 SUDS, LID, BMPs, WSUD and more – The evolution and application of terminology surrounding urban drainage. *Urban Water Journal* **12** (7), 525–542. https://doi.org/10.1080/1573062X.2014.916314.

Fowdar, H., Payne, E., Schang, C., Zhang, K., Deletic, A. & McCarthy, D. 2021 How well do stormwater green infrastructure respond to changing climatic conditions? *Journal of Hydrology* **603**, 126887. https://doi.org/10.1016/j.jhydrol.2021.126887.

Fu, B., Liu, Y., Lü, Y., He, C., Zeng, Y. & Wu, B. 2011 Assessing the soil erosion control service of ecosystems change in the Loess Plateau of China. *Ecological Complexity* **8** (4), 284–293. https://doi.org/10.1016/j.ecocom.2011.07.003.

Fu, X., Hopton, M. E. & Wang, X. 2021 Assessment of green infrastructure performance through an urban resilience lens. *Journal of Cleaner Production* **289**, 125146. https://doi.org/10.1016/j.jclepro.2020.125146.

Gordon, B. L., Quesnel, K. J., Abs, R. & Ajami, N. K. 2018 A case-study based framework for assessing the multi-sector performance of green infrastructure. *Journal of Environmental Management* **223**, 371–384. https://doi.org/10.1016/j.jenvman.2018.06.029.

Grabowski, Z. J., McPhearson, T., Matsler, A. M., Groffman, P. & Pickett, S. T. 2022 What is green infrastructure? A study of definitions in US city planning. *Frontiers in Ecology and the Environment* **20** (3), 152–160. https://doi.org/10.1002/fee.2445.

Griffiths, J., Chan, F. K. S., Shao, M., Zhu, F. & Higgitt, D. L. 2020 Interpretation and application of Sponge City guidelines in China. *Philosophical Transactions of the Royal Society A* **378** (2168), 20190222. https://doi.org/10.1098/rsta.2019.0222.

Grol, R., Dalhuijsen, J., Thomas, S., Rutten, G. & Mokkink, H. 1998 Attributes of clinical guidelines that influence use of guidelines in general practice: Observational study. *BMJ* **317** (7162), 858–861. https://doi.org/10.1136/bmj.317.7162.858.

Hansen, R. & Pauleit, S. 2014 From multifunctionality to multiple ecosystem services? A conceptual framework for multifunctionality in green infrastructure planning for urban areas. *Ambio* **43**, 516–529. https://doi.org/10.1007/s13280-014-0510-2.

Healy, R. W. & Cook, P. G. 2002 Using groundwater levels to estimate recharge. *Hydrogeology Journal* **10** (1), 91–109. https://doi.org/10.1007/s10040-001-0178-0.

Hooftman, D. A., Kimberley, A., Cousins, S. A., Bueno, S. S., Honnay, O., Krickl, P., Plue, J., Poschlod, P., Traveset, A. & Bullock, J. M. 2023 Could green infrastructure supplement ecosystem service provision from semi-natural grasslands? *Journal of Environmental Management* **328**, 116952. https://doi.org/10.1016/j.jenvman.2022.116952.

Jarvis, A., Varma, A. & Ram, J. 2011 *Assessing Green Jobs Potential in Developing Countries: A Practitioner's Guide*. International Labour Office, Geneva. ISBN 978-92-2-124571-1.

Kabisch, N. & Haase, D. 2013 Green spaces of European cities revisited for 1990–2006. *Landscape and Urban Planning* **110**, 113–122. https://doi.org/10.1016/j.landurbplan.2012.10.017.

Langeveld, J. G., Cherqui, F., Tscheikner-Gratl, F., Muthanna, T. M., Juarez, M. F.-D., Leitão, J. P., Roghani, B., Kerres, K., do Céu Almeida, M., Werey, C. & Rulleau, B. 2022 Asset management for blue-green infrastructures: A scoping review. *Blue-Green Systems* **4** (2), 272–290. https://doi.org/10.2166/bgs.2022.019.

Lin, W., Chen, Q., Jiang, M., Zhang, X., Liu, Z., Tao, J., Wu, L., Xu, S., Kang, Y. & Zeng, Q. 2019 The effect of green space behaviour and per capita area in small urban green spaces on psychophysiological responses. *Landscape and Urban Planning* **192**, 103637. https://doi.org/10.1016/j.landurbplan.2019.103637.

Liu, L. & Jensen, M. B. 2018 Green infrastructure for sustainable urban water management: Practices of five forerunner cities. *Cities* **74**, 126–133. https://doi.org/10.1016/j.cities.2017.11.013.

Lopes, M. N. & Camanho, A. S. 2013 Public green space use and consequences on urban vitality: An assessment of European cities. *Social Indicators Research* **113** (3), 751–767. https://doi.org/10.1007/s11205-012-0106-9.

Lucas, W. C. & Sample, D. J. 2015 Reducing combined sewer overflows by using outlet controls for green stormwater infrastructure: Case study in Richmond, Virginia. *Journal of Hydrology* **520**, 473–488. https://doi.org/10.1016/j.jhydrol.2014.10.029.

Lundgren, L. 2021 *Metal Removal Efficiency of Five Filter Media Intended for Use in Road Stormwater Treatment Facilities. Degree Project in Environmental Engineering and Sustainable Infrastructure*. KTH Royal Institute of Technology, Stockholm, Sweden.

Maimaitiyiming, M., Ghulam, A., Tiyip, T., Pla, F., Latorre-Carmona, P., Halik, Ü., Sawut, M. & Caetano, M. 2014 Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaptation. *ISPRS Journal of Photogrammetry and Remote Sensing* **89**, 59–66. https://doi.org/10.1016/j.isprsjprs.2013.12.010.

Manso, M., Teotónio, I., Silva, C. M. & Cruz, C. O. 2021 Green roof and green wall benefits and costs: A review of the quantitative evidence. *Renewable and Sustainable Energy Reviews* **135**, 110111. https://doi.org/10.1016/j.rser.2020.110111.

Matsler, A. M., Meerow, S., Mell, I. C. & Pavao-Zuckerman, M. A. 2021 A 'green chameleon': Exploring the many disciplinary definitions, goals, and forms of 'green infrastructure'. *Landscape and Urban Planning* **214**, 104145. https://doi.org/10.1016/j.landurbplan.2021.104145.

Mintz, D. 2009 *Technical Assistance Document for the Reporting of Daily Air Quality-the Air Quality Index (AQI)*. Tech. Research Triangle Park, US Environmental Protection Agency, Durham, NC.

Moor, H., Eggers, J., Fabritius, H., Forsell, N., Henckel, L., Bradter, U., Mazziotta, A., Nordén, J. & Snäll, T. 2022 Rebuilding green infrastructure in boreal production forest given future global wood demand. *Journal of Applied Ecology*. https://doi.org/10.1111/1365-2664.14175.

NSW Department of Planning and Environment 2022 Interim Framework for Valuing Green Infrastructure and Public Spaces – Technical Appendices for Recommended Approaches, New South Wales, Australia.

Ommer, J., Bucchignani, E., Leo, L. S., Kalas, M., Vranić, S., Debele, S., Kumar, P., Cloke, H. L. & Di Sabatino, S. 2022 Quantifying co-benefits and disbenefits of nature-based solutions targeting disaster risk reduction. *International Journal of Disaster Risk Reduction* **75**, 102966. https://doi.org/10.1016/j.ijdrr.2022.102966.

Orsini, F., Gasperi, D., Marchetti, L., Piovene, C., Draghetti, S., Ramazzotti, S., Bazzocchi, G. & Gianquinto, G. 2014 Exploring the production capacity of rooftop gardens (RTGs) in urban agriculture: The potential impact on food and nutrition security, biodiversity and other ecosystem services in the city of Bologna. *Food Security* **6** (6), 781–792. https://doi.org/10.1007/s12571-014-0389-6.

Othus-Gault, S. 2021 Measuring carbon sequestration. *National Earth Science Teachers Association* **37** (1), 33–36.

Oxford English Dictionary. Oxford University Press, Oxford 2023 Retrieved from https://www.oed.com/

Pakzad, P. & Osmond, P. 2016 Developing a sustainability indicator set for measuring green infrastructure performance. *Procedia-social and Behavioral Sciences* **216**, 68–79. https://doi.org/10.1016/j.sbspro.2015.12.009.

Peschardt, K. K., Schipperijn, J. & Stigsdotter, U. K. 2012 Use of small public urban green spaces (SPUGS). *Urban Forestry & Urban Greening* **11** (3), 235–244. https://doi.org/10.1016/j.ufug.2012.04.002.

Phillips, D., Jamwal, P., Lindquist, M. & Gronewold, A. 2022 Assessing catchment-scale performance of in-stream provisional green infrastructure interventions for dry weather flows. *Landscape and Urban Planning* **226**, 104448. https://doi.org/10.1016/j.landurbplan.2022.104448.

Radcliffe, J. C. 2019 History of water sensitive urban design/low impact development adoption in Australia and internationally. In: *Approaches to Water Sensitive Urban Design* (Sharma, A. K., Gardner, T. & Begbie, D., eds.). Elsevier, Amsterdam, pp. 1–24. https://doi.org/10.1016/B978-0-12-812843-5.00001-0.

Raymond, C. M., Breil, M., Nita, M., Kabisch, N., de Bel, M., Enzi, V., Frantzeskaki, N., Geneletti, G., Lovinger, L. & Cardinaletti, M. 2017 *An Impact Evaluation Framework to Support Planning and Evaluation of Nature-Based Solutions Projects*. Report prepared by the EKLIPSE Expert Working Group on Nature-Based Solutions to Promote Climate Resilience in Urban Areas. Centre for Ecology and Hydrology.

Ricci, A., Guasco, M., Caboni, F., Orlanno, M., Giachetta, A. & Repetto, M. P. 2022 Impact of surrounding environments and vegetation on wind comfort assessment of a new tower with vertical green park. *Building and Environment* **207**, 108409. https://doi.org/10.1016/j.buildenv.2021.108409.

Rosenzweig, C., Solecki, W. & Slosberg, R. 2006 Mitigating New York City's heat island with urban forestry, living roofs, and light surfaces. A report to the New York State Energy Research and Development Authority, 1–5.

Sabia, G., Petta, L., Avolio, F. & Caporossi, E. 2020 Energy saving in wastewater treatment plants: A methodology based on common key performance indicators for the evaluation of plant energy performance, classification and benchmarking. *Energy Conversion and Management* **220**, 113067. https://doi.org/10.1016/j.enconman.2020.113067.

Sánchez, R., José, E. S., Fermoso, J., Gómez, S., González, M., 2022 Types of nature-based solutions for water, their functions and suitability. In: *Regional Perspectives of Nature-Based Solutions for Water: Benefits and Challenges. Applied*

Environmental Science and Engineering for a Sustainable Future (Pachova, N., Velasco, P., Torrens, A. & Jegatheesan, V., eds). Springer, Cham. https://doi.org/10.1007/978-3-031-18412-3_2.

Seo, S., Choi, S., Kim, K., Kim, S. M. & Park, S. M. 2019 Association between urban green space and the risk of cardiovascular disease: A longitudinal study in seven Korean metropolitan areas. *Environment International* **125**, 51–57. https://doi.org/10.1016/j.envint.2019.01.038.

Silva, C., Cardoso, M. A., Rosa, M. J., Alegre, H., Ugarelli, R., Bosco, C., Raspati, G., Azrague, K., Bruaset, S., Damman, S., Koop, S., Munaretto, S., Silva da Conceição, M. M., Gomes, C., Rosell, L. F., Schmuck, A., Strehl, C. & Mohan Doss, P. 2023 *Final Version of the Water-Smartness Assessment Framework – PDF (Deliverable 6.3)*. https://b-watersmart.eu/download/final-version-of-the-water-smartness-assessment-framework-pdf-deliverable-6-3/#

Sohn, W., Bae, J. & Newman, G. 2021 Green infrastructure for coastal flood protection: The longitudinal impacts of green infrastructure patterns on flood damage. *Applied Geography* **135**, 102565. https://doi.org/10.1016/j.apgeog.2021.102565.

Sun, J., Cheshmehzangi, A. & Wang, S. 2020 Green infrastructure practice and a sustainability key performance indicators framework for neighbourhood-level construction of sponge city programme. *Journal of Environmental Protection* **11** (2), 82–109. https://doi.org/10.4236/jep.2020.112007.

Tang, J., Wang, W., Feng, J., Yang, L., Ruan, T. & Xu, Y. 2021 Urban green infrastructure features influence the type and chemical composition of soil dissolved organic matter. *Science of the Total Environment* **764**, 144240. https://doi.org/10.1016/j.scitotenv.2020.144240.

Threlfall, C. G., Mata, L., Mackie, J. A., Hahs, A. K., Stork, N. E., Williams, N. S. & Livesley, S. J. 2017 Increasing biodiversity in urban green spaces through simple vegetation interventions. *Journal of Applied Ecology* **54** (6), 1874–1883. https://doi.org/10.1111/1365-2664.12876.

Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J. & James, P. 2007 Promoting ecosystem and human health in urban areas using green infrastructure: A literature review. *Landscape and Urban Planning* **81** (3), 167–178. https://doi.org/10.1016/j.landurbplan.2007.02.001.

Ugarelli, R., Damman, S., Koop, S., Alegre, H., Cardoso, M. A., Almeida, M., Oliveira, R., Vildåsen, S. S., Raspati, G., Barrero, M. A., Termes, M., Mateo, M. R., Schmuck, A., Strehl, C. & Lekawska-Andrinopoulou, L. 2021 *What is Water Smartness and How to Assess It (Deliverable 6.1)*. Available from: https://b-watersmart.eu/download/what-is-water-smartness-and-how-to-asses-it-d6-1/#

United Nations Information Portal on Multilateral Environmental Agreements (InforMEA). Green Infrastructure. Available from: https://www.informea.org/en/terms/green-infrastructure (accessed March 2024)

USEPA 2024 *What is Green Infrastructure?* Available from: https://www.epa.gov/green-infrastructure/what-green-infrastructure (accessed March 2024)

Vaissière, A. C., Levrel, H., Hily, C. & Le Guyader, D. 2013 Selecting ecological indicators to compare maintenance costs related to the compensation of damaged ecosystem services. *Ecological Indicators* **29**, 255–269. https://doi.org/10.1016/j.ecolind.2013.01.003.

Verheijen, F., Jeffery, S., Bastos, A. C., Van der Velde, M. & Diafas, I. 2010 Biochar application to soils. A critical scientific review of effects on soil properties, processes, and functions. *EUR* **24099** (162), 2183–2207. https://doi.org/10.2788/472.

Wang, W., Zhang, L. & Li, J. 2020 *Assessment Standard for Sponge City Effects*. IWA Publishing, London. https://doi.org/10.2166/9781789060553.

Wenzler, I. 2005 Development of an asset management strategy for a network utility company: Lessons from a dynamic business simulation approach. *Simulation & Gaming* **36** (1), 75–90. https://doi.org/10.1177/1046878104272668.

Wolsink, M. 2016 Environmental education excursions and proximity to urban green space–densification in a 'compact city'. *Environmental Education Research* **22** (7), 1049–1071. https://doi.org/10.1080/13504622.2015.1077504.

Xiao, Q., McPherson, E. G., Simpson, J. R. & Ustin, S. L. 2007 Hydrologic processes at the urban residential scale. *Hydrological Processes: An International Journal* **21** (16), 2174–2188. https://doi.org/10.1002/hyp.6482.

Xie, L. M. & Wang, W. Y. 2012 A brief introduction to appraisal of guidelines for research and evaluation II. *Zhong xi yi jie he xue bao Journal of Chinese Integrative Medicine* **10** (2), 160–165. https://doi.org/10.3736/jcim20120206.

Xu, C., Tang, T., Jia, H., Xu, M., Xu, T., Liu, Z., Long, Y. & Zhang, R. 2019 Benefits of coupled green and grey infrastructure systems: Evidence based on analytic hierarchy process and life cycle costing. *Resources, Conservation and Recycling* **151**, 104478. https://doi.org/10.1016/j.resconrec.2019.104478.

Young, R. F. & McPherson, E. G. 2013 Governing metropolitan green infrastructure in the United States. *Landscape and Urban Planning* **109** (1), 67–75. https://doi.org/10.1016/j.landurbplan.2012.09.004.

Yu, J. 2020 Exploring the role of healthy green spaces, psychological resilience, attitude, brand attachment, and price reasonableness in increasing hotel guest retention. *International Journal of Environmental Research and Public Health* **17** (1), 133. https://doi.org/10.3390/ijerph17010133.

Ziter, C. D., Pedersen, E. J., Kucharik, C. J. & Turner, M. G. 2019 Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer. *Proceedings of the National Academy of Sciences* **116** (15), 7575–7580. https://doi.org/10.1073/pnas.181756111.

Zuniga-Teran, A. A., Gerlak, A. K., Mayer, B., Evans, T. P. & Lansey, K. E. 2020 Urban resilience and green infrastructure systems: Towards a multidimensional evaluation. *Current Opinion in Environmental Sustainability* **44**, 42–47. https://doi.org/10.1016/j.cosust.2020.05.001.

First received 27 October 2023; accepted in revised form 18 April 2024. Available online 31 May 2024