Put et al. Systematic Reviews (2025) 14:254
https://doi.org/10.1186/s13643-025-02939-4

Systematic Reviews

PROTOCOL

Open Access

Determinants and outcomes of advanced practice nurses' leadership behaviours: an AI-aided mixed-methods systematic review protocol

Vincent Put[1], Hanne Kindermans[1], Ann Van Hecke[2,3], Greta G. Cummings[4] and Ellen Vlaeyen[1,5*]

Abstract

Background Advanced practice nurses play a vital role in healthcare innovation, delivering high-quality care and improving patient outcomes. Leadership is a core competency of advanced practice nurses, empowering them to drive systemic improvements and foster collaboration. However, these master-level educated nurses often encounter challenges in assuming leadership roles, including limited recognition and competing demands on their time. The growing volume of healthcare-related research, combined with the lack of a comprehensive evidence base on the determinants and outcomes of their leadership behaviours, complicates the development of effective programmes. This protocol outlines a systematic approach to addressing these challenges, using an AI tool to efficiently manage the expanding evidence base and provide a detailed understanding of the factors influencing advanced practice nurses' leadership behaviours.

Methods This protocol follows the PRISMA-P 2015 guidelines to outline a systematic review investigating the determinants and outcomes of advanced practice nurses' leadership behaviours. It employs the SPIDER tool for eligibility criteria, encompassing studies that explore advanced practice nursing leadership behaviours and their determinants and outcomes. Eligible studies include quantitative, qualitative and mixed-methods research, focusing on advanced practice nursing roles. The protocol also outlines a workflow for AI-aided title and abstract screening using ASReview LAB, incorporating multi-phase human validation to ensure accuracy and reliability. Data synthesis will utilise narrative synthesis for quantitative data and meta-aggregation for qualitative findings, integrating results through narrative weaving.

Discussion This protocol addresses a critical gap in nursing research by systematically exploring the determinants influencing advanced practice nurses' leadership behaviours and their outcomes. It provides evidence to inform the development of tailored programmes aimed at empowering advanced practice nurses to maximise their leadership potential. Additionally, the protocol demonstrates how AI tools can enhance systematic review efficiency while maintaining methodological rigour. The findings will not only contribute to advancing nursing practice but also highlight the transformative potential of AI in research synthesis, ensuring timely and robust evidence generation amidst the expanding volume of healthcare-related research.

Systematic review registration PROSPERO CRD42025644174.

*Correspondence:
Ellen Vlaeyen
ellen.vlaeyen@uhasselt.be
Full list of author information is available at the end of the article

Keywords Leadership, Advanced practice nursing, Behavioural research, Systematic review, Artificial intelligence, Machine learning

Introduction

The emergence of advanced practice nursing roles has been described as one of the most transformative developments in the profession during the twentieth century [1]. This recognition is supported by their ability to consistently deliver high-quality care and enhance patient outcomes across healthcare settings [2–4]. The International Council of Nurses [5] defines an advanced practice nurse (APN) as 'a generalist or specialized nurse who has acquired, through additional graduate education (minimum of a master's degree), the expert knowledge base, complex decision-making skills and clinical competencies for advanced nursing practice, the characteristics of which are shaped by the context in which they are credentialed to practice.' (p. 6). Beyond their clinical expertise, APNs demonstrate the ability to drive innovation and foster improvements in healthcare systems, with leadership emerging as a key competency critical to these efforts [6].

Despite its recognised importance, research indicates that APNs dedicate limited time to leadership activities [7], citing perceived gaps in the necessary competencies for their execution [8]. Large clinical caseloads have also been shown to impair APNs' ability to enact leadership [9, 10], as the demands of direct patient care leave little time for broader organisational or system-level engagement. Consequently, APNs may be under-recognised as formal leaders [11], even though leadership is an expected component of advanced practice nursing [6]. Moreover, the limited evidence base on advanced practice leadership complicates policymakers' efforts in preparing, developing and evaluating advanced practice nursing roles [12]. When the skills of these professionals are not understood, their potential may remain underutilised, resulting in missed opportunities that could otherwise benefit patients, healthcare providers and the broader health system [11]. To address these challenges, targeted efforts are needed to support APNs in cultivating behaviours that enhance their impact as healthcare leaders.

One potential solution lies in developing a behaviour-change programme, defined as a coordinated set of activities designed to change specific behavioural patterns [13]. Given that programmes based solely on a planner's intuition are most likely to prove ineffective [14], programme development should be grounded in theory and evidence that captures the distinct mechanisms of action involved in achieving behavioural change [15]. As such, prominent behaviour-change frameworks highlight the critical first step of analysing the relationships, whether observed or hypothesised, between the target behaviour, its influencing factors and its consequences [13, 16, 17].

Behavioural influences, commonly referred to as determinants in behaviour-change literature, encompass the internal factors (i.e. biological and psychosocial) and environmental conditions that shape an individual's behaviour [18]. The identification of relevant determinants is crucial, as it allows programme planners to pinpoint what needs to be targeted and select appropriate behaviour-change techniques to address them effectively [19]. Behavioural consequences, or outcomes, are the changes at individual, community or systemic levels resulting from programme activities and corresponding behaviours [20]. By identifying these outcomes, planners can establish measurable indicators of change that offer insights into whether the programme has effectively met its goals, allowing comparisons against agreed-upon standards or benchmarks [21]. More specifically, they provide a way to measure quality and bring attention to the often-overlooked contributions of APNs' leadership in advancing nursing and the broader healthcare system [22]. As such, a systematic approach to developing a behaviour-change programme for advanced practice nursing leadership should, at a minimum, include an examination of both the determinants and outcomes of this behaviour [15]. While barriers and enablers representing some of these determinants have recently been systematically reviewed in a study of cross-sectional research published from 2015 onwards [10], a comprehensive synthesis of both determinants and outcomes across the scientific literature, which includes qualitative evidence, remains lacking.

Systematic reviews have become a cornerstone of policy decision-making, with policymakers increasingly recognising their value in supporting evidence-based decisions [23]. However, the growing volume of unstructured scientific literature has rendered traditional evidence synthesis increasingly time-consuming and resource-intensive, threatening its practicality [24]. For instance, an analysis of the International Prospective Register of Systematic Reviews (PROSPERO) found that it takes a median of 65.8 weeks from protocol registration to publication [25], and as many as 23% of systematic reviews may require updating within just 2 years of publishing [26]. Policymakers, particularly in dynamic fields like healthcare, view the delays between review completion and when its findings are needed as a significant obstacle [27]. To address this challenge, researchers are exploring alternative methods, which include artificial intelligence (AI) tools that have the potential to significantly reduce workloads while preserving reviewer oversight and methodological integrity [28].

To support the screening process, the review will utilise ASReview LAB, an open-source and freely available platform that enhances efficiency through active learning based on AI [29]. ASReview LAB ranks records by predicting their relevance using features derived from previously screened studies, allowing the model to prioritise which records are presented to the reviewer. The developers validated this approach in simulation studies on multiple existing review datasets, identifying 95% of eligible studies after screening just 8 to 33% of records. These findings highlight its potential to support more efficient evidence synthesis without compromising the identification of relevant literature. This approach will be used to aid in title and abstract screening to answer the following research questions: 'What are the determinants of APNs' leadership behaviours?' and 'What are the outcomes of APNs' leadership behaviours?'.

Methods

The review is registered on PROSPERO with the registration number CRD42025644174. The conduct of the review will be guided by the JBI Manual for Evidence Synthesis [30]. To promote comprehensive and transparent reporting, this protocol follows the reporting guidelines set by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Protocols statement [31] (see Additional file 1). The review itself will follow the PRISMA 2020 reporting guidelines [32].

Eligibility criteria

The eligibility criteria were established using the SPIDER (Sample, Phenomenon of Interest, Design, Evaluation and Research type) search strategy tool [33].

The sample of interest consists of nurses working in advanced practice roles in any healthcare-related setting (e.g. hospitals, nursing homes, general practitioners' offices). Initial study selection based on sample will focus on the APN title, which most often encompasses the internationally recognised roles of nurse practitioner and clinical nurse specialist. Given the significant variability in APNs' role descriptions [34], studies using a generic advanced practice nursing title (e.g. advanced practice role, advanced practice provider, advanced clinical practitioner) will also be considered eligible for review. If the title in an individual study is unclear or ambiguous, the sample will be assessed based on the International Council of Nurses' [5] definition of an APN, which requires that the nurse provides direct patient or client care requiring an expert knowledge base or advanced skills, and holds a master's degree in a health-related discipline.

For design and research type, the review will include peer-reviewed primary research that used quantitative, qualitative or mixed-methods study designs. Non-empirical studies such as abstracts, book chapters, conference proceedings, editorials, meta-analyses, opinion papers, protocols and reviews will be excluded.

Regarding the phenomenon of interest, studies will be eligible for review if leadership is explicitly stated as a study aim. Given the absence of a definitive definition of APNs' leadership [35], it may be examined either as a leadership style or as any behaviour aligning with Northouse's [36] broad conceptualisation, which states that leadership is a process whereby an individual influences a group of individuals to achieve a common goal. In cases where a study investigates APNs' tasks or competencies without explicitly identifying leadership as an aim, inclusion will be determined based on whether the abstract explicitly mentions leadership as part of their competencies or responsibilities. Studies focusing on leadership in managerial or hierarchical roles (e.g. head nurses, directors of nursing, healthcare administrators) will be excluded from review.

The evaluation of the review will focus on the experienced or measured relationships between APNs' leadership behaviours and the corresponding determinants or outcomes of those behaviours. Determinants refer to all internal mechanisms, such as biological (e.g. age, sex, genetic predispositions) and psychosocial factors (e.g. skills, knowledge, beliefs, habits, emotions), and external environmental mechanisms (e.g. social influences, workplace culture, organisational policies, physical surroundings) that have the potential to influence behaviour [18]. For outcomes, a distinction is made between personal outcomes (e.g. job satisfaction, psychological well-being, organisational commitment, turnover intention), patient-level outcomes (e.g. clinical outcomes, patient satisfaction, adherence to treatment) and systemic outcomes (e.g. job performance, team effectiveness, profitability).

Search strategy

The electronic databases that were selected for the review are CINAHL (via EBSCOhost), Embase (via Elsevier), PubMed (via NCBI, including MEDLINE), Scopus (via Elsevier) and Web of Science Core

Collection (webofscience.com; SCI-EXPANDED, SSCI, AHCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI). The search strategy was developed in collaboration with two information specialists from the Knowledge Centre for Health Ghent and was structured around the SPIDER framework, focusing on the 'Sample' (i.e. advanced practice nurses) and the 'Phenomenon of Interest' (i.e. leadership). As the review does not target specific methodological approaches, the 'Design' and 'Research type' components were excluded, in order to capture articles that might not contain reference to these elements. The 'Evaluation' component was also omitted, as predefining specific determinants or outcomes would introduce an arbitrary filter and potentially exclude relevant evidence. As pertaining to the concept of leadership, we have deliberately included the term 'management' to account for its frequent interchangeable use with leadership in the literature [37], thereby minimising the risk of omitting relevant studies that examine leadership-related concepts described through the lens of management terminology. The search string was initially designed for the PubMed search engine and then adapted for use in the other aforementioned databases. The original search string and its translations are presented in Table 1 and Additional file 2, respectively. The search string incorporates conceptually synonymous free-text terms in titles, abstracts and keywords alongside controlled (medical) vocabulary (e.g. CINAHL subject headings, Emtree terms, MeSH). Databases will be searched from their respective inception to the date of search. Additionally, backwards reference cross-checking of all included full-text studies will be conducted to identify further relevant records. Only studies published in Dutch or English will be considered for inclusion. No additional search for gray literature will be performed.

AI-aided study screening and selection

Following the search, the resulting studies will be uploaded to the reference management software EndNote for data storage and an initial round of deduplication [38]. Given the heightened importance of deduplication in AI-aided screening, where multiple decisions regarding the same record could disproportionately influence the AI's classification algorithm [39], and the relatively low specificity, sensitivity and accuracy of EndNote in detecting duplicates [40], a second, more precise round will be conducted using the web-based software DedupEndNote [41]. The DedupEndNote software consistently outperforms EndNote for deduplication, achieving greater sensitivity (i.e. 93.1–98.6% vs. 51.2–74.4%), specificity (i.e. 99.9–100% vs. 80.3–99.8%) and accuracy (i.e. 94.5–99.5% vs. 72.0–93.6%) across multiple datasets [41].

Title and abstract screening will be facilitated by the free and open-source research software ASReview LAB, which provides reviewers with a machine-learning-aided pipeline that integrates an active learning cycle to systematically estimate the most relevant records for human screening [29]. Active learning is a subfield within AI in which a model (i.e. a configuration of various algorithms and methods) can choose the data points from which it

Table 1 Search strategy for the PubMed (via NCBI, including MEDLINE) database

SPIDER[a] term	Concept	Search strategy
Sample (S)	Advanced practice nurse	(advance practice role*[tiab] AND nurs*[tiab]) OR "advanced practice nurs*"[tiab] OR (APN[tiab] AND nurs*[tiab]) OR (ANP[tiab] AND nurs*[tiab]) OR "nurse practitioner*"[tiab] OR (NP[tiab] AND nurs*[tiab]) OR "advanced nurs*"[tiab] OR "advanced clinical nurs*"[tiab] OR "clinical nurse specialist*"[tiab] OR (CNS[tiab] AND nurs*[tiab]) OR "clinical nursing specialist*"[tiab] OR "advanced practice registered nurs*"[tiab] OR "Advanced Practice Nursing"[MeSH Terms] OR "Nurse Practitioners"[MeSH Terms]
Phenomenon of interest (P)	Leadership	"lead*"[tiab] OR "manag*"[tiab] OR "Leadership"[MeSH Terms]
S + P	Combination	((advance practice role*[tiab] AND nurs*[tiab]) OR "advanced practice nurs*"[tiab] OR (APN[tiab] AND nurs*[tiab]) OR (ANP[tiab] AND nurs*[tiab]) OR "nurse practitioner*"[tiab] OR (NP[tiab] AND nurs*[tiab]) OR "advanced nurs*"[tiab] OR "advanced clinical nurs*"[tiab] OR "clinical nurse specialist*"[tiab] OR (CNS[tiab] AND nurs*[tiab]) OR "clinical nursing specialist*"[tiab] OR "advanced practice registered nurs*"[tiab] OR "Advanced Practice Nursing"[MeSH Terms] OR "Nurse Practitioners"[MeSH Terms]) AND ("lead*"[tiab] OR "manag*"[tiab] OR "Leadership"[MeSH Terms])

[a] Sample, Phenomenon of Interest, Design, Evaluation, Research type

learns by querying an oracle (i.e. a human user or another source of information) to assign labels to previously unlabelled documents, thereby minimising the resource investment in acquiring labelled data [42]. In ASReview LAB, these labels are binary (i.e. relevant or irrelevant) and are assigned to a record's title and abstract [29].

First, three independent reviewers will initiate the screening process by labelling 100 randomly selected studies from the database according to the eligibility criteria. Following this preparatory screening, the three reviewers will meet to discuss any discrepant labels based on their individual interpretations of the eligibility criteria to establish a shared understanding for the remainder of the study screening. This process also allows for the calculation of a crude estimate of the total number of relevant studies in the dataset by multiplying the fraction of relevant studies (i.e. the number of relevant studies in the preparatory screening step divided by the total number of studies in the preparatory screening step) by the total number of studies [43]. The resulting estimate will later inform the stopping criterion for the title and abstract screening process.

Second, the complete study database will be uploaded into a locally run instance of ASReview LAB, where it will be screened by reviewer 1. The labelling consensus from the preparatory screening step will serve as training data for the active learning model, providing an initial basis to determine a record's relevance. In ASReview LAB, this training data must include at least one study labelled as relevant and one labelled as irrelevant [29]. The model for this step will use the following specifications: term frequency-inverse document frequency (TF-IDF) as the feature extractor, Naive Bayes as the classifier, dynamic resampling and a maximum query strategy. These are the default settings within ASReview LAB [44], selected for their consistently strong performance on benchmark tests across multiple datasets [45]. The feature extraction technique converts text into a structured format (e.g. TF-IDF, Sentence BERT, Doc2Vec) that the classifier can process. The classifier, in turn, is the algorithm responsible for calculating relevance scores based on these extracted features (e.g. Naive Bayes, logistic regression, random forest, neural network). A balancing strategy is used to mitigate overfitting to the majority class (i.e. irrelevant records in systematic reviews), thereby enhancing the model's ability to generalise to new, unseen data. Lastly, the query strategy determines which documents are presented to the oracle for labelling after the model computes the relevance scores. The available options are clustering, which selects representative samples from different groups of similar documents; maximum certainty-based, where the model prioritises the most relevant documents based on its highest confidence scores; and uncertainty-based, which selects documents where the model's confidence is lowest. The random strategy selects documents entirely at random, ignoring the model's predictions, while the mixed strategy combines 95% maximum certainty-based with either 5% uncertainty or 5% random selection, balancing confident predictions with exploration of less certain or random areas of the dataset. Next, the active learning cycle begins as the software presents one record at a time for the reviewer to label as relevant or irrelevant. Based on this feedback, the AI model is retrained and the process repeats, improving its predictions with each new label until the reviewer chooses to stop or a predefined stopping criterion is reached [44]. A stopping criterion is used to determine when to end the active learning process in screening software, weighing the cost of continued screening against the likelihood of overlooking relevant records. It will be set based on the SAFE procedure: (1) a selection of predefined key papers has been presented to the human reviewer by the AI model, (2) at least twice the estimate of the total number of relevant studies has been screened, (3) a minimum of 10% of the dataset has been screened and (4) no study has been labelled as relevant in the last 100 records [43]. Before screening commences, the research team will compile a list of landmark studies to be used for this process. Once reviewer 1 has met the stopping criterion, reviewers 2 and 3 will each independently screen a randomly selected but distinct half of the studies suggested to reviewer 1 by the AI model. Discrepancies in labels after this step will be resolved through discussion by the research team.

Third, reviewer 1 will conduct a second round of AI-aided screening of titles and abstracts using a more complex model: Sentence Bidirectional Encoder Representations from Transformers (SBERT) as the feature extractor, random forest as the classifier, dynamic resampling and a mixed query strategy (95% maximum-certainty and 5% uncertainty) [44]. It is recommended to begin screening with a simpler model and transition to a more computationally demanding one once more labelled data is available, as these models require more training data to perform optimally [46]. Therefore, the research team's consensus on inclusions and exclusions from the previous step will be fully utilised as training data for this more complex model. The output of this screening step will then again be passed to reviewers 2 and 3, with each independently screening a randomly selected but distinct half of the studies suggested by the model. Discrepancies in labels after this step will be resolved by the research team.

Last, the three reviewers will conduct a manual full-text screening against the eligibility criteria, with reviewer 1 screening all studies and reviewers 2 and 3 each independently screening a randomly selected but distinct half of the studies. The research team will resolve any remaining discrepancies in inclusion decisions in a final review.

Study quality assessment

A sensitivity analysis will be performed on all included full-text studies, utilising the approach introduced by Vandervelde and colleagues [47]. This analysis combines the methodological quality assessment (i.e. low, moderate, high) and the relevance to the research questions (i.e. low, moderate, high). Studies with low ratings in both categories, or a low rating in one category and moderate in the other, will be classified as low overall. Studies with moderate ratings in both categories, or a combination of moderate and high ratings, will be classified as moderate overall. Only studies rated high in both categories will receive a high overall rating. Although studies will not be excluded based on their sensitivity analysis score, those with a low overall rating will be interpreted with caution during synthesis and integration.

Methodological quality assessment will be conducted using the Mixed-Methods Appraisal Tool (MMAT) as described by Hong and colleagues [48]. Each category of study design includes five distinct methodological criteria, assessed as 'yes' (i.e. criterion met), 'no' (i.e. criterion not met) or 'can't tell'. Studies will be rated based on the number of criteria met: 'low' if 0–1 criteria are met, 'moderate' if 2–3 are met and 'high' if 4–5 are met. For mixed-methods studies, each methodological component will be assessed separately. Reviewer 1 will rate the methodological quality of all included studies, while reviewers 2 and 3 will each independently evaluate a randomly selected but distinct half of the studies. Any discrepancies will be resolved through discussion among the research team to ensure consensus.

Relevance to the research questions will be evaluated using criteria developed by the research team prior to the review process. These criteria will be designed to assess the extent to which each study contributes meaningful and applicable insights into the determinants and outcomes of APNs' leadership behaviours. Relevance to the research questions will be categorised as 'low,' 'moderate,' or 'high,' based on the finalised criteria. Reviewer 1 will assess all studies for relevance, while reviewers 2 and 3 will each independently assess a randomly selected but distinct half of the studies. Any discrepancies will be resolved through discussion among the research team to ensure consensus.

Data extraction

Reviewer 1 will independently extract data from all included full-text studies into a customised Microsoft Excel spreadsheet, while reviewers 2 and 3 will independently extract data from a randomly selected but distinct half of the studies. To ensure the spreadsheet captures all information relevant to the research questions, it will be piloted using a sample of six studies, randomly selecting two quantitative, two qualitative and two mixed-methods studies. Any modifications to the spreadsheet will be agreed upon by consensus among the research team and documented in the systematic review publication. Data will be extracted across five categories: (1) publication details (e.g. study title, name of first author, publication year, journal of publication), (2) research design (e.g. research type, research methodology), (3) sample characteristics (e.g. gender and/or sex distribution, mean and/or median age, sample size, professional APN title, educational level, country of practice, care setting(s) of practice, nursing specialisation), (4) phenomenon of interest (e.g. APNs' leadership behaviour(s) and/or style(s)) and (5) study findings (e.g. experienced and/or measured relationship(s) between APNs' leadership behaviour(s) and corresponding determinants and/or outcomes). The complete list of data extraction points is presented in tabular form in Additional file 3. Encountered inconsistencies or disagreements during data extraction will be resolved through collective discussion among the research team. In cases of missing or unclear data, corresponding authors will be contacted up to two times by email: an initial request will be sent, followed by a second request after four weeks if no response is received. The data will be considered unobtainable if no response is received after the second request.

Data synthesis and integration

Data will be synthesised using a results-based convergent synthesis design [49]. Qualitative and quantitative studies, including the individual methodologies within mixed-methods studies, will first be synthesised and presented separately. If these methodologies are not clearly distinguishable in the text, the corresponding authors will be contacted to provide this information. The results from both syntheses will then be integrated with a third synthesis, reconsidering the evidence in the context of their combined findings [49].

Given the broad scope of the review, which seeks to include all potential determinants and outcomes of APNs' leadership behaviours reported in the literature, in addition to the utilised wide-ranging conceptualisation of these behaviours, a high degree of methodological and intervention heterogeneity is anticipated. The sheer diversity of variables under investigation renders statistical pooling through methods like meta-analysis unfeasible [50], necessitating the use of narrative synthesis to comprehensively analyse the quantitative data. This synthesis will be presented as a narrative elaboration of identified patterns, complemented by study characteristics and statistical data displayed in tables or other pertinent visual formats [50].

Qualitative data will be synthesised using the meta-aggregation method [51]. Extracted study findings, defined as verbatim excerpts of the study authors' analytical interpretations of their data, will be grouped into categories comprising two or more similar findings. Each finding will be accompanied by an illustration, such as a direct quotation from a study participant, a fieldwork observation or another supporting data point from the study [51]. As each finding is extracted, it will be assigned a level of plausibility based on the reviewers' assessment of the congruency between the finding and its accompanying illustration. The three levels are (1) unequivocal (i.e. findings accompanied by an illustration that is beyond reasonable doubt and not open to challenge), (2) equivocal (i.e. findings with an illustration that lacks an explicit association and is therefore open to challenge) and (3) unsupported (i.e. findings not supported by the data). Unequivocal findings, ranked the highest, will be prioritised in the synthesis, followed by equivocal findings. Unsupported findings will be excluded from the synthesis [52]. For each created category, the research team will draw up an explanatory statement summarising the shared theme or concept underlying the group of similar findings. Lastly, synthesised findings will be created by combining two or more groups of categories, also accompanied by an explanatory statement. These statements are drawn up to be suggestive, characteristic, representative, symbolic or emblematic of the evidence being synthesised [51, 52]. While such statements are traditionally used to inform policy or practice, in our review, they will guide further research, particularly in the development of a programme. In both cases, the purpose of these indicative statements remains consistent: to distil the essence of the evidence into actionable insights, making this approach equally suitable for shaping research directions or informing policy. If textual pooling is not feasible, a narrative synthesis will be conducted as an alternative to meta-aggregation, following a structure similar to that used for synthesising the quantitative data.

The integration of quantitative and qualitative data will occur at the interpretation and reporting level through narrative weaving, where both data types are organised on a concept-by-concept basis [53]. A joint display will be used to enhance the visual interpretation of the integrated findings, with the specific framework to be determined during synthesis in such a manner that it aligns with the emerging results.

Discussion

Leadership is widely integrated as a central competency in prominent frameworks for advanced practice nursing, as it enables APNs to effectively influence patient care, foster collaboration and drive improvement in healthcare systems [54]. Despite its recognised importance, a comprehensive synthesis of the factors that shape and influence APNs' leadership behaviours has not been conducted to date. Understanding these relationships is vital for identifying the factors influencing how APNs adopt and perform leadership roles, particularly as they navigate the dual demands of direct patient care and broader organisational responsibilities. The planned review seeks to address this critical gap in nursing literature by systematically gathering and analysing the scientific evidence on the determinants and outcomes of APNs' leadership behaviours. The findings of this review will help the development of approaches that enable APNs to fully realise their leadership potential within the complex environments in which they practice.

Complementing this effort, the current protocol also aims to advance systematic review methodology within nursing literature by introducing AI-aided screening of titles and abstracts using ASReview LAB. This tool has been shown to reduce screening workloads by an average of 60.02% [55], effectively tackling one of the most resource-intensive aspects of traditional systematic reviews. Moreover, by reducing the volume of records requiring manual screening, AI tools can help decrease the potential for labelling errors, which can reach up to 21% among experienced reviewers and 58% among inexperienced ones [56]. As the volume of nursing-related research continues to expand rapidly, incorporating AI into the screening process offers a robust, future-oriented solution, ensuring that systematic reviews remain rigorous and capable of meeting the demands of an ever-growing evidence base. Additionally, the open-source nature of ASReview LAB aligns with the principles of transparency and reproducibility in systematic reviews, allowing researchers to evaluate its underlying functionality and adapt it to their specific needs.

The use of AI in systematic reviews, while innovative, is not without limitations. Although it offers significant potential to reduce screening workload, this efficiency may come at the cost of missing up to 5% of relevant studies [57]. This limitation arises from the model's reliance on iterative learning, which prioritises only a subset of records for screening rather than reviewing the entire dataset. Additionally, it is important to acknowledge that AI-assisted approaches may not completely eliminate biases in study selection, as they still rely on human judgment for inclusion or exclusion decisions. These concerns align with broader discussions in the literature about both the potential benefits and inherent risks of using AI in evidence synthesis [30]. As technological advancements in this space continue to outpace established methodological frameworks, their adoption must be approached with caution and critical reflection. To address these challenges in our protocol, we have incorporated multiple strategies to optimise the synergy between AI tools and human reviewers. Every study suggested by the AI model is reviewed by at least two individuals to ensure thorough assessment, while studies rejected by the AI undergo cross-validation using two distinct AI models. This layered approach minimises the risk of missing critical evidence while ensuring a balanced and reliable review process.

Despite its current limitations, integrating AI into the screening process represents a transformative step forward in addressing the challenges posed by the expanding volume of nursing-related research. By enhancing efficiency, reducing the burden of screening and maintaining methodological rigour, AI provides a scalable and future-oriented solution for systematic reviews. This integration ensures that evidence synthesis remains feasible and responsive to the growing demands of evidence-based nursing practice while paving the way for continued methodological advancements in research synthesis across disciplines.

Abbreviations

APN	Advanced practice nurse
AI	Artificial intelligence
MMAT	Mixed-Methods Appraisal Tool
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	International Prospective Register of Systematic Reviews
SBERT	Sentence Bidirectional Encoder Representations from Transformers
SPIDER	Sample, Phenomenon of Interest, Design, Evaluation, Research Type
TF-IDF	Term Frequency-Inverse Document Frequency

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13643-025-02939-4.

Additional file 1: Table S1. PRISMA-P1 2015 checklist: recommended items to address in a systematic review protocol.

Additional file 2: Table S1. Search strategies for the CINAHL, Embase, Scopus, and Web of Sciences Core Collection databases.

Additional file 3: Data extraction points in tabular form.

Acknowledgements
The authors wish to thank Nele Pauwels and Muguet Koobasi, information specialists at the Knowledge Centre for Health Ghent (library of Ghent University Hospital and the faculties of Medicine & Health Sciences & Pharmaceutical Sciences), for their help in developing the review search strategy.

Authors' contributions
All authors contributed to the study conceptualisation. All authors contributed to the design of the methodology, including the search strategy. VP drafted and revised the manuscript based on the critical appraisal of HK, AVH, GGC and EV. EV supervised the process of protocol development. All authors read and approved the final manuscript.

Funding
This study was supported by the Special Research Fund (reference: BOF24OWB12) of UHasselt (Hasselt University). The funder had no role in the study's conceptualisation, design, methodology development or manuscript preparation.

Data availability
The search strings described in this protocol are included in the published article and its supplementary information file. For the systematic review publication, the search strings will be submitted to searchRxiv (https://searchrxiv.org/), where a Digital Object Identifier (DOI) will be generated.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Faculty of Medicine and Life Sciences, Research Group Healthcare & Ethics, UHasselt, Agoralaan, Diepenbeek 3590, Belgium. [2]Faculty of Medicine and Health Sciences, Department of Public Health and Primary Care, University Center for Nursing and Midwifery, Ghent University, Ghent 9000, Belgium. [3]Center of Nursing Excellence, Ghent University Hospital, Ghent 9000, Belgium. [4]Faculty of Nursing, College of Health Sciences, University of Alberta, Edmonton, AB, Canada. [5]Faculty of Medicine, Department of Public Health and Primary Care, Academic Centre for Nursing and Midwifery, KU Leuven, Kapucijnenvoer 7, bus 7001, Leuven 3000, Belgium.

Received: 12 February 2025 Accepted: 27 August 2025
Published online: 30 December 2025

References
1. Oddsdóttir EJ, Sveinsdóttir H. The content of the work of clinical nurse specialists described by use of daily activity diaries. J Clin Nurs. 2011;20(9–10):1393–404. https://doi.org/10.1111/j.1365-2702.2010.03652.x.
2. Donald F, Martin-Misener R, Carter N, Donald EE, Kaasalainen S, Wickson-Griffiths A, et al. A systematic review of the effectiveness of advanced practice nurses in long-term care. J Adv Nurs. 2013;69(10):2148–61. https://doi.org/10.1111/jan.12140.
3. Swan M, Ferguson S, Chang A, Larson E, Smaldone A. Quality of primary care by advanced practice nurses: a systematic review. Int J Qual Health Care. 2015;27(5):396–404. https://doi.org/10.1093/intqhc/mzv054.
4. Woo BF, Lee JX, Tam WW. The impact of the advanced practice nursing role on quality of care, clinical outcomes, patient satisfaction, and cost in the emergency and critical care settings: a systematic review. Hum Resour Health. 2017;15(1):63. https://doi.org/10.1186/s12960-017-0237-9.
5. International Council of Nurses. Guidelines on advanced practice nursing 2020. Geneva: International Council of Nurses; 2020. 39 p. Available from https://www.icn.ch/resources/publications-and-reports/guidelines-advanced-practice-nursing-2020.
6. Dowling M, Beauchesne M, Farrelly F, Murphy K. Advanced practice nursing: a concept analysis. Int J Nurs Pract. 2013;19(2):131–40. https://doi.org/10.1111/ijn.12050.
7. Goemaes R, Lernout E, Goossens S, Decoene E, Verhaeghe S, Beeckman D, et al. Time use of advanced practice nurses in hospitals: a

cross-sectional study. J Adv Nurs. 2019;75(12):3588–601. https://doi.org/10.1111/jan.14198.
8. Van Hecke A, Goemaes R, Verhaeghe S, Beyers W, Decoene E, Beeckman D. Leadership in nursing and midwifery: activities and associated competencies of advanced practice nurses and midwives. J Nurs Manag. 2019;27(6):1261–74. https://doi.org/10.1111/jonm.12808.
9. Elliott N, Begley C, Sheaf G, Higgins A. Barriers and enablers to advanced practitioners' ability to enact their leadership role: a scoping review. Int J Nurs Stud. 2016;60:24–45. https://doi.org/10.1016/j.ijnurstu.2016.03.001.
10. Fosah R, Llahana S. Barriers and enablers to leadership in advanced practice nursing: a systematic review. Int Nurs Rev. 2025;72(2):e70034. https://doi.org/10.1111/inr.70034.
11. Lamb A, Martin-Misener R, Bryant-Lukosius D, Latimer M. Describing the leadership capabilities of advanced practice nurses using a qualitative descriptive study. Nurs Open. 2018;5(3):400–13. https://doi.org/10.1002/nop2.150.
12. Elliott N, Higgins A, Begley C, Lalor J, Sheerin F, Coyne I, et al. The identification of clinical and professional leadership activities of advanced practitioners: findings from the specialist clinical and advanced practitioner evaluation study in Ireland. J Adv Nurs. 2013;69(5):1037–50. https://doi.org/10.1111/j.1365-2648.2012.06090.x.
13. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci. 2011;6:42. https://doi.org/10.1186/1748-5908-6-42.
14. Kass NE. An ethics framework for public health. Am J Public Health. 2001;91(11):1776–82. https://doi.org/10.2105/AJPH.91.11.1776.
15. Bartholomew LK, Mullen PD. Five roles for using theory and evidence in the design and testing of behavior change interventions. J Public Health Dent. 2011;71:20–33. https://doi.org/10.1111/j.1752-7325.2011.00223.x.
16. Bartholomew Eldredge LK, Markham CM, Ruiter RA, Fernández ME, Kok G, Parcel GS. Planning health promotion programs: an intervention mapping approach. 4th ed. San Francisco (CA): Jossey-Bass; 2016. p. 678.
17. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:2061. https://doi.org/10.1136/bmj.n2061.
18. Kok G, Peters LW, Ruiter RA. Planning theory- and evidence-based behavior change interventions: a conceptual review of the intervention mapping protocol. Psicol Reflex Crit. 2017;30(1):19. https://doi.org/10.1186/s41155-017-0072-x.
19. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Appl Psychol. 2008;57(4):660–80. https://doi.org/10.1111/j.1464-0597.2008.00341.x.
20. Newcomer KE, Hatry HP, Wholey JS. Meeting the need for practical evaluation approaches: an introduction. In: Wholey JS, Hatry HP, Newcomer KE, editors. Handbook of practical program evaluation. 2nd ed. San Francisco (CA): Jossey-Bass; 2004. p. xxxiii-xliv.
21. Griffiths P, Jones S, Maben J, Murrells T. State of the art metrics for nursing: a rapid appraisal. London: National Nursing Research Unit; 2008. p. 38.
22. Elliott N, Begley C, Kleinpell R, Higgins A. The development of leadership outcome-indicators evaluating the contribution of clinical specialists and advanced practitioners to health care: a secondary analysis. J Adv Nurs. 2014;70(5):1078–93. https://doi.org/10.1111/jan.12262.
23. Fox DM. Evidence and health policy: using and regulating systematic reviews. Am J Public Health. 2017;107(1):88–92. https://doi.org/10.2105/AJPH.2016.303485.
24. Bornmann L, Haunschild R, Mutz R. Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. Humanit Soc Sci Commun. 2021;8:224. https://doi.org/10.1057/s41599-021-00903-w.
25. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017;7(2):e012545. https://doi.org/10.1136/bmjopen-2016-012545.
26. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007;147(4):224–33. https://doi.org/10.7326/0003-4819-147-4-200708210-00179.
27. Tricco AC, Cardoso R, Thomas SM, Motiwala S, Sullivan S, Kealey MR, et al. Barriers and facilitators to uptake of systematic reviews by policy makers and health care managers: a scoping review. Implement Sci. 2016;11:4. https://doi.org/10.1186/s13012-016-0370-1.
28. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Syst Rev. 2019;8:163. https://doi.org/10.1186/s13643-019-1074-9.
29. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nat Mach Intell. 2021;3:125–33. https://doi.org/10.1038/s42256-020-00287-7.
30. Aromataris E, Lockwood E, Porritt K, Pilla B, Jordan Z, editors. JBI manual for evidence synthesis. JBI; 2024. https://doi.org/10.46658/JBIMES-24-01.
31. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1. https://doi.org/10.1186/2046-4053-4-1.
32. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev. 2021;10:89. https://doi.org/10.1186/s13643-021-01626-4.
33. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res. 2012;22(10):1435–43. https://doi.org/10.1177/1049732312452938.
34. De Raeve P, Davidson PM, Bergs J, Patch M, Jack SM, Castro-Ayala A, et al. Advanced practice nursing in Europe—results from a pan-European survey of 35 countries. J Adv Nurs. 2024;80(1):377–86. https://doi.org/10.1111/jan.15775.
35. Heinen M, van Oostveen C, Peters J, Vermeulen H, Huis A. An integrative review of leadership competencies and attributes in advanced nursing practice. J Adv Nurs. 2019;75(11):2378–92. https://doi.org/10.1111/jan.14092.
36. Northouse PG. Leadership: theory and practice. 7th ed. Thousand Oaks (CA): SAGE Publications; 2016. p. 494.
37. Algathani A. Are leadership and management different? A review. J Manag Policies Pract. 2014;2(3):71–82. https://doi.org/10.15640/jmpp.v2n3a4.
38. The EndNote Team. EndNote [software]. Version 21.5. Philadelphia (PA): Clarivate; 2013.
39. van Dijk SH, Brusse-Keizer MG, Bucsán CC, van der Palen J, Doggen CJ, Lenferink A. Artificial intelligence in systematic reviews: promising when appropriately used. BMJ Open. 2023;13(7):e072254. https://doi.org/10.1136/bmjopen-2023-072254.
40. McKeown S, Mir ZM. Considerations for conducting systematic reviews: evaluating the performance of different methods for de-duplicating references. Syst Rev. 2021;10(1):38. https://doi.org/10.1186/s13643-021-01583-y.
41. Lobbestael G. DedupEndNote [software]. Version 1.0.1. Available from https://github.com/globbestael/DedupEndNote.
42. Settles B. Active learning literature survey. Madison (WI): University of Wisconsin–Madison; 2010. 65 p. Computer Sciences Technical Report 1648. Available from https://burrsettles.com/pub/settles.activelearning.pdf.
43. Boetje J, van de Schoot R. The SAFE procedure: a practical stopping heuristic for active learning-based screening in systematic reviews and meta-analyses. Syst Rev. 2024;13:81. https://doi.org/10.1186/s13643-024-02502-7.
44. de Bruin J, Ferdinands G, Harkema A, de Boer J, van den Brand S, Ma Y, van de Schoot R. ASReview LAB software documentation. Version v1.3. https://doi.org/10.5281/zenodo.10066693.
45. Ferdinands G, Schram R, de Bruin J, Bagheri A, Oberski DL, Tummers L, et al. Performance of active learning models for screening prioritization in systematic reviews: a simulation study into the average time to discover relevant records. Syst Rev. 2023;12:100. https://doi.org/10.1186/s13643-023-02257-7.
46. Teijema JJ, Hofstee L, Brouwer M, de Bruin J, Ferdinands G, de Boer J, et al. Active learning-based systematic reviewing using switching classification models: the case of the onset, maintenance, and relapse of depressive disorders. Front Res Metr Anal. 2023;8:1178181. https://doi.org/10.3389/frma.2023.1178181.
47. Vandervelde S, Van den Bosch N, Vlaeyen E, Dierckx de Casterlé B, Flamaing J, Belaen G, Tuand K, Vandendriessche T, Milisen K. Determinants influencing the implementation of multifactorial falls risk assessment

and multidomain interventions in community-dwelling older people: a systematic review. Age Ageing. 2024;53(7):afae123. https://doi.org/10.1093/ageing/afae123.
48. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. Educ Inf. 2018;34(4):285–91. https://doi.org/10.3233/EFI-180221.
49. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev. 2017;6:61. https://doi.org/10.1186/s13643-017-0454-2.
50. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme. Lancaster: University of Lancaster; 2006. p. 92.
51. Lockwood C, Porritt K, Munn Z, Rittenmeyer L, Salmond S, Bjerrum M, Loveday H, Carrier J, Stannard D. Systematic reviews of qualitative evidence. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI manual for evidence synthesis. 2024 edition. JBI; 2024. p. 58–112. Available from https://synthesismanual.jbi.global.
52. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. Int J Evid Based Healthc. 2015;13(3):179–87. https://doi.org/10.1097/XEB.0000000000000062.
53. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs—principles and practices. Health Serv Res. 2013;48:2134–56. https://doi.org/10.1111/1475-6773.12117.
54. Arslanian-Engoren C. Conceptualizations of advanced practice nursing. In: Tracy MF, O'Grady ET, Phillips SJ, editors. Hamric & Hanson's advanced practice nursing: an integrative approach. 7th ed. St. Louis (MO): Elsevier; 2023. p. 31–73.
55. Chan YT, Abad JE, Dibart S, Kernitsky JR. Assessing the article screening efficiency of artificial intelligence for systematic reviews. J Dent. 2024;149:105259. https://doi.org/10.1016/j.jdent.2024.105259.
56. Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. BMC Med Res Methodol. 2019;19:132. https://doi.org/10.1186/s12874-019-0782-0.
57. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev. 2015;4:5. https://doi.org/10.1186/2046-4053-4-5.

Publisher's Note