Journal of Clinical Nursing

WILEY

Journal of Clinical Nursing

RESEARCH METHODOLOGY: DISCUSSION PAPER – METHODOLOGY OPEN ACCESS

Learnings, Challenges and Processes of Prioritizing Initial Programme Theories on Contextually Appropriate Nurse Staffing Models With a Multinational Team: A Realist Methodology Case Example

Kaitlyn Tate[1] | Andrew Booth[2] | Gillian Harvey[3] | Tony Ryan[4] | Inge Wolbers[5] | Pieterbas Lalleman[6] | Candice Oster[3] | Rachel Flynn[7] | Tatiana Penconek[1] | Matthias Hoben[1,8] | Tim Schultz[9] | Greta G. Cummings[1] | on behalf of the ReSoNANCE (Realist Synthesis of Nursing in Australia Netherlands Canada and England) group

[1]College of Health Sciences, Faculty of Nursing, University of Alberta, Edmonton, Alberta, Canada | [2]School of Medicine and Population Health, University of Sheffield, Sheffield, UK | [3]College of Nursing and Health Sciences, Flinders University, Adelaide, South Australia, Australia | [4]School of Allied Health, Pharmacy, Nursing and Midwifery, University of Sheffield, Sheffield, UK | [5]University of Applied Sciences Utrecht, Utrecht, the Netherlands | [6]Fontys University of Applied Science, Eindhoven, the Netherlands | [7]School of Nursing and Midwifery, University College Cork, Cork, Ireland | [8]School of Health Policy and Management, Faculty of Health, York University, Toronto, Ontario, Canada | [9]College of Medicine and Public Health, Flinders University, Adelaide, South Australia, Australia

Correspondence: Greta G. Cummings (gretac@ualberta.ca)

Received: 17 September 2025 | **Revised:** 27 June 2026 | **Accepted:** 7 July 2026

Keywords: context | initial programme theories | multinational | nurse staffing | realist methodology | realist review | ReSoNANCE

ABSTRACT

Aim: To describe the processes, consensus-building activities, learning and challenges of prioritizing initial programme theories on appropriate nurse staffing models across a multinational research team of academics, trainees and knowledge users from four countries.

Design: Realist review involving literature review, multinational stakeholder engagement and prioritization of programme theories. Realist reviews explore what works, for whom and in what contexts, enabling the development and testing of initial programme theories.

Data Sources: Data sources included (1) a literature review to identify existing theories on optimal nurse staffing models, (2) input from methodological and health system decision makers experts to identify context-mechanism-outcome configurations and prioritize initial programme theories and (3) feedback from health decision-makers and nursing staff through focus groups to augment and prioritize a final set of initial programme theories for subsequent testing.

Outcomes: We summarize the key learnings we gained regarding: (1) effectively communicating the value and unique aspects of realist methodology to team members with varying levels of expertise; (2) achieving a compromise between rigour and relevance in the design of the review itself in working within 'social laboratories' and (3) the logistics of conducting a multinational realist investigation: navigating working remotely, differences in time zones and differences in context and language.

Conclusion: Prioritizing initial programme theories on nurse staffing through a multinational realist review is both feasible and valuable. Clear communication, balancing rigour with relevance and managing logistical challenges are essential to producing context-sensitive and evidence-informed staffing models.

Relevance to Clinical Practice: Developing evidence-informed staffing models requires local context consideration and stakeholder collaboration. This study offers (a) practical insights for nurse leaders and managers on balancing rigour and relevance, (b) effective team communication and (c) navigating logistical challenges in multinational teams to prioritize programme theories for clinical translation to improve patient care and staff wellbeing.

Reporting Methods: Consolidated criteria for reporting qualitative research (COREQ).

Patient or Public Contribution: Health care decision makers, care managers, researchers and academicians were involved.

What This Paper Adds to the Wider Global Clinical Community

- Demonstrates how a multinational realist approach can generate context-sensitive and evidence-informed nurse staffing models with global relevance.
- Provides practical strategies for balancing methodological rigour with local context and for communicating realist methods across teams with varied expertise.
- Offers transferable lessons on managing logistical challenges in international research, supporting collaboration to improve safe staffing and workforce sustainability worldwide.

1 | Introduction

Globally, the need for improved and effective nurse staffing across various health settings is being recognized (World Health Organization 2024). As the largest workforce in health systems, nurse staffing affects patient outcomes as well as health system costs and productivity (Aiken et al. 2014; Sherenian et al. 2013). The nursing workforce has faced many challenges over the past 20 years, including nursing shortages and high rates of burnout and staff turnover (Tamata and Mohammadnezhad 2023; van Oostveen et al. 2015). Workforce issues were exacerbated by the COVID-19 pandemic, which created additional challenges for nurse staffing due to understaffing and further increases in burnout and staff turnover (Boniol et al. 2022; International Council of Nurses 2023; Roush 2020).

Several approaches to determining appropriate numbers, grades and skill mix of nurses include nurse-to-patient ratios, patient acuity and historical staffing. These approaches may be used on their own or in combination to inform an organization's nurse staffing. Despite a large amount of research produced on this topic, there is little conclusive evidence from traditional systematic reviews about the effectiveness of various staffing models and a lack of synthesis on *why* a model may be effective or not, limiting the contextual understanding of nurse staffing (Aiken et al. 2014; Sherenian et al. 2013).

Realist reviews seek to answer questions of *what works, for whom, in what circumstances, how and why* by reconceptualizing the causal model in which intervention X leads to a given outcome (O). Realism asks what contexts (C) and mechanisms (M) allow X to lead to each outcome (O), so the X-O model becomes a CMO model (Pawson et al. 2005). This CMO model allows the realist review process to be explanatory and sensitive to context. This helps when a topic is under-theorized and research exploring the effectiveness of an intervention or programme is inconclusive, as in the case of nurse staffing models. The CMO model supports the development of an initial programme theory, which is needed to explicitly frame how we expect an intervention to work, for whom and importantly, in which particular contexts. Having outlined early assumptions about the program's mechanisms and contexts, evidence from the synthesis is used to test, confirm, refute or refine aspects of the initial theory. The aim of any realist synthesis is to develop a robust and evidence-based understanding of how, why and within which contexts (conditions) the program works.

To address complex problems pragmatically in the real world, focusing on context is essential to avoid unintended consequences and differential outcomes. Contexts include social, economic and political structures; healthcare settings; organizational settings or social conditions; program participants; and geographical and historical context (Pawson et al. 2005). Mechanisms refer to responses which are not directly observable or the 'underlying entities, processes or [social] structures which operate in particular contexts of interest'. (p. 368) (Astbury and Leeuw 2010). Outcomes are products or results of mechanisms that are triggered in certain contexts (Long et al. 2022).

Guidelines exist for conducting realist reviews. Advances have been made in regard to more detailed and transparent reporting to enhance rigour and potential replication of realist approaches. Specifically, detailed descriptions and recommendations are available around literature review and synthesis approaches to identify, elicit and refine initial programme theory, in addition to explanations on how to think about and identify latent mechanisms (Flynn et al. 2020; Smeets et al. 2022). Some guidance exists for determining thresholds of relevance, rigour and richness of description for quality appraisal in realist reviews, albeit these appraisals are interpretive in nature (Dada et al. 2023). There are limited resources available that unpack these iterative and interpretive decision-making processes during initial programme theory development across various contexts within a large, multi-national team (Jagosh et al. 2011). Furthermore, given the large amount of evidence on the topic of nurse staffing and the number and varieties of contexts that could be explored, prioritizing initial programme theory development and testing is necessary to ensure that the work is feasible as well as relevant to local contexts in which it will be applied. Yet, little guidance on how to approach these processes is available, especially among a diverse, multinational team.

1.1 | Research Context

The overarching aim of our research was to conduct a multiphase realist review with a multi-national team to examine in which contexts and through what mechanisms nurse staffing models produce client, professional and health system outcomes (Pawson et al. 2005; Wong et al. 2013, 2012). Here, we report on the approaches taken during phase one of this realist review, a key phase where cross-country collaboration and decision-making were paramount. Our research team included Canadian, Australian, Dutch and British health service researchers, methodologists, senior knowledge users, academic leaders and trainees. Each team had two senior researchers with expertise in the substantive content, one team member with methodological expertise and/or experience, one senior knowledge user and at least one trainee. These roles were not always mutually exclusive. Many team members also had clinical backgrounds as Registered Nurses. This four-country team builds on previously developed relationships and harnesses critical expertise. The richness of our international team, encompassing many similarities and dissimilarities, ensured we had knowledge of a range of healthcare contexts to consider and compare, supporting comparative consideration across diverse healthcare contexts. Each member brought a unique perspective to elucidate hidden mechanisms and how they operate in these contexts. In this sense, each team was thought of as its own social 'laboratory' in which we brought together various stakeholders and engaged in iterative learning cycles to address a health system challenge in full consideration of real-life settings (Marschalek et al. 2022; Timmermans et al. 2020). A multinational team was required, not just to interpret context simplistically in terms of temporal and geographical variation, but to capture diverse data that elicits intra-country variations and render a truly realist interpretation of context.

1.2 | Nurse Staffing Context and Research

The existing empirical landscape on nurse staffing models is heavily dominated by quantitative research examining the structural components of care delivery. For over two decades, research syntheses, including systematic reviews, have primarily focused on evaluating the direct relationships between isolated staffing inputs, such as registered nurse-to-patient ratios, skill mix proportions and nurse substitution strategies and various clinical or workforce outcomes (Butler et al. 2019; Dall'Ora et al. 2022; Tate et al. 2024). Within this body of work, researchers have explored whether maximizing registered nurse (RN) hours directly minimizes adverse events, whether substituting RNs with less qualified or unregulated nursing assistive personnel yields operational efficiencies and other variations on nurse staffing models (Bridges et al. 2019; Butler et al. 2019). However, major systematic evaluations frequently conclude that the overall data regarding non-fatal patient outcomes, staff-patient interactions and optimal model configurations remain highly inconclusive, structurally weak, or equivocal across different health systems (Bridges et al. 2019; Butler et al. 2019; Dall'Ora et al. 2022).

Significant gaps persist in the literature due to methodological limitations and a historical lack of contextualization of not just findings, but the research questions themselves. Methodologically, the majority of historical evidence linking staffing configurations to care quality relies on cross-sectional designs, and even emerging longitudinal data struggles to establish definitive causal pathways due to residual confounding (Dall'Ora et al. 2022). Existing literature features a distinct geographic and resource bias with most studies conducted within high-income nations (Imam et al. 2022). Crucially, traditional systematic reviews have historically treated staffing inputs as fixed formulas, largely overlooking how nonlinear, shifting operational realities, such as variations in local ward-level culture, complexity and acuity of patients/clients, or the complex supervisory burdens placed on RNs during nurse substitution, confound outcomes (Bridges et al. 2019; Tate et al. 2024). By failing to account for how macro-level regulatory environments and micro-level clinical climates interact, standard evaluations have left a theoretical gap regarding how, for whom and in what specific circumstances these diverse staffing arrangements actually function (Tate et al. 2024).

1.3 | Aim

To describe the processes, consensus-building activities, learning and challenges of prioritizing initial programme theories on appropriate nurse staffing models across a multinational research team of academics, trainees and knowledge users from four countries. In this paper, we report on the aspects of this work for which we optimized the nature of a multinational team, explored contextual differences and experienced challenges for which we will expand on lessons learned.

2 | Data Sources

Here, we report on methodological approaches taken as a multinational team during phase one of a three-phase realist review which encompassed multiple iterative and overlapping stages. During phase one, we:

a. conducted a review of the literature to identify existing theories on optimal nurse staffing models.

b. consulted with our methodological and health system decision-maker experts to identify CMO configurations and from them develop IPTs.

c. solicited stakeholder feedback from a broad group of health system decision-makers and nursing staff through focus groups to augment and prioritize a final set of IPTs for subsequent theory testing.

2.1 | Literature Review

Standard review processes were used in terms of developing search strategies, searching databases and screening procedures. Inclusion criteria for the literature review were papers that examined a nurse staffing model (either with or without other healthcare professionals) conceptualized, developed, implemented, or evaluated in any of the following health settings (in-patient and out-patient settings including: acute care settings

[ED and in hospital], primary care settings, community care settings, palliative care settings, transitional or rehabilitation care settings and continuing care settings), in high income countries based on World Bank income groups (https://data.worldbank.org/income-level/high-income). Key areas of multinational collaboration and decision-making for this phase are highlighted in the procedures described. These inclusion and exclusion criteria were discussed as a team and applied independently by pairs of senior researchers and trainees to search results based on title and abstracts. Pairs were selected from different regions to enhance team building and ensure diverse perspectives were discussed. Trainees completed screening procedures after an initial 10% of abstracts were screened by paired senior researchers and trainees. A lead methodologist (AB) reviewed the included sample and triaged papers for inclusion in subsequent steps. Specifically, we employed the 3Rs framework for triaging, in which theoretical statements identified in the included papers had to be relevant to our research questions, rigorous and rich in content (i.e., clearly linking context, mechanism and outcome) (Jagosh et al. 2011). Senior research team members, methodologists and team members who conducted the screening reviewed the triaged papers. Any discrepant views or clarifications were managed via email correspondence. Resulting Context, Mechanism and Outcome configurations were extracted from nine articles and prioritized for further consolidation and refinement. Detailed processes on these steps are reported elsewhere (Tate et al. Forthcoming).

2.2 | CMOc Consolidation and Prioritization

2.2.1 | Initial Prioritization Working Meetings

A list of extracted CMOcs was sent to all team members prior to an initial working meeting so that team members could rank CMOcs of relevance to their context. We conducted two separate 2-h working meetings to expand on and prioritize CMOcs. We accommodated various time zones, ensuring a lead methodologist was represented within each working group. Each country team aimed to have Principal Applicants, a methodologist, a decision-maker and two or three substantive nurse/workforce experts at their respective working meetings. One group included members from the UK and Canada, and the other group included members from Australia and The Netherlands. We agreed that up to 15 CMOcs (from the nine included papers) could be feasibly managed within the team and would represent the diversity of key ideas presented across the CMOcs. Facilitators presented the results (rankings of CMOcs) from the pre-meeting exercise to each group, who then discussed, clarified and expanded on CMOcs of interest. For these working meetings, we had a shared goal and aimed for similar processes across the two subgroups; however, one subgroup did more asynchronous ranking and consolidation of CMOcs as there seemed to be higher agreement across team members in pre-meeting work on the CMOcs, which resulted in more focused discussion during the working meeting and an output of 'rough' initial programme theories. The other subgroup did not find the pre-meeting ranking exercise helpful and engaged in more discussion-based identification of themes or ideas for the focus of initial programme theories from the long list of CMOcs. In both groups, CMOcs were consolidated and expansion of CMOcs included hypothesizing mechanisms and clarifying contexts through consensus discussion among researchers with substantive expertise and experiential knowledge as nurses as well as knowledge users (health system decision-makers) on our team. Both outputs were shared asynchronously, discussed via email and then compared to the other subgroup output and linked more explicitly to source articles in order to bridge these outputs into a mutually agreed upon list of 'rough' initial programme theories. Ultimately, we decided that nursing workforce experts on our team would lead sensemaking of these CMOcs through facilitated discussion. We agreed we would aim for a mix of common/different CMOcs, seeking not to achieve replicability across countries/research teams but, rather, to examine and contextualize CMOcs most relevant to each country team. For example, in The Netherlands, there is no explicit legislation around nurse-to-patient ratios, so a CMOc related to this was not found to be relevant and prioritized for this country team. In this initial prioritization meeting among our research team, we agreed on CMOc consolidation and suggested additions to the CMOcs were added in coloured text, with synthesized CMOcs subsequently finalized by team members through consensus discussion. This resulted in the development of 'rough' initial programme theories. A set of initial programme theories, judged as potentially relevant to all country teams, was confirmed through asynchronous email communication.

Stakeholder Engagement. Between May 2022 and September 2022, we held one to two focus groups in each country with key nursing stakeholders (e.g., executive directors, policy makers, unions, healthcare managers at low, middle and upper levels) to elicit their understanding of the theories that underpin staffing models. To recruit participants, information letters were e-mailed to relevant stakeholder mailing lists, which, depending on the country, was facilitated by senior health decision makers on our research team.

Each country team selected the most appropriate stakeholders based on their context. Individuals or groups external to the team referred and identified relevant participants, so long as these potential participants had a role with an organization for which operational approval was obtained and they contacted the research staff to participate. Health decision makers on our research team sent an information letter providing information on the study as well as contact information of the Principal Applicant and research staff of each country via staff mailing lists. Interested participants contacted the Principal Applicant or research staff/coordinator for that country. Potential participants' contact information was kept in a password-protected document accessible only to the study team.

2.2.2 | Preparing for and Conducting Focus Groups

In consultation with team methodologists, we developed a focus group guide with main questions and probing questions (see File S1), informed by the RAMESES II Project Realist Interview and Realist Evaluation Interviewing- A 'Starter Set' of Questions (Project 2017). Country teams could select IPTs according to their relevance to their region and stakeholder preference. Specific procedures, developed and shared with each

country team, covered the logistics for focus groups as guidelines including (a) identification and recruitment of participants, (b) arranging, conducting and recording focus groups.

All country teams conducted focus groups to further prioritize and refine initial programme theories. Focus groups were conducted virtually via Zoom or Microsoft Teams. The interviewer obtained informed oral consent from participants. First, we provided a brief overview of the included initial programme theories for discussion. Then, we asked participants to select those on which to focus discussion (or, if preferred, to exclude any not considered important). These focus groups were suited to (a) examining the integrity of a specific staffing theory, (b) comparing rival theories, (c) reviewing the same theory in comparative settings and (d) reviewing official expectations against actual practice.

In total, 10 focus groups were held across all country teams with a total of 43 participants. Each focus group consisted of 4–6 participants. In Canada, we recruited a total of 12 included unit managers and senior leaders (e.g., directors, program managers) separately. Program areas for participants in senior leadership roles included medicine, neurology, neonatal intensive care, women's health, acute care, emergency, public health, personal care homes and rural sites across Alberta. Unit managers had experience in family medicine & surgery units, COVID-19 designated units, critical care, emergency, paediatrics and ambulatory clinics. In the UK, 10 participants comprised of a blend of senior nurses (decision making) and Modern Matrons (Modern Matrons are senior nursing leaders responsible for planning and delivering patient care). All participants were based in acute care settings. In the Netherlands, eight participants comprised of district nurses, nurse managers and experts on staffing from acute care hospital setting, home care and long-term care facilities. In Australia, we recruited 13 participants including chief nurses, executive directors and union employees with responsibility for workforce in healthcare.

2.3 | Ethical Considerations

We obtained ethical approval from the respective ethical review boards in each country, specifically for the involvement of stakeholders in focus group discussions (Ethical approval #Pro00100425). Verbal informed consent was obtained from all participants, and their contact information was stored in a password-protected document accessible only to the study team. No demographic information was collected from the participants. Focus groups were conducted virtually using encrypted Zoom or Microsoft Outlook. Audio recordings were securely stored in a password-protected folder within the health repository data environment.

2.4 | Analysis, Refinement and Prioritization of Initial Programme Theories

Through facilitated discussion around the meaningfulness of each theory, participants provided information on if and how the initial programme theories on nurse staffing models were relevant to their particular contexts and how they did or did not produce patient, staff and health system outcomes. We asked participants to rank or prioritize initial programme theories for further testing. Each country team recorded and transcribed focus group discussions. Each country team analysed their focus group findings and held meetings in which they summarized their focus group findings to include the following: main points from the discussion, feedback on the initial programme theories discussed, priorities identified and anything participants felt was missing from the initial programme theories.

Each country team provided a summary of findings to all teams prior to a large team discussion. We held an all-country team meeting to discuss the focus group summaries, as well as identify and reach consensus on initial programme theories to prioritise for theory testing. We compared how each country's focus group participants prioritized initial programme theories. Four initial programme theories were identified as prioritized by participants for three or four country teams. We determined that to capitalize on our resources and diverse contexts, each country would lead the theory testing (through empirical literature) for one of four initial programme theories. For the Australian team, one initial programme theory on mandated nurse–patient ratios was identified as the most relevant for their context based on their stakeholder engagement and was not as emphasised as important for other countries. The remaining theories were discussed among team members and assigned based on relative importance to each team and the level of refinement and augmentation of the initial programme theory based on focus group findings. For instance, for one initial programme theory, the Canadian team had developed a rival theory based on study findings (wherein having sufficient Registered Nurse staffing with a complement of nursing assistants produced positive outcomes in one theory, but negative outcomes in another) and therefore this theory was assigned to the Canadian team. Ultimately, consensus discussion produced the final prioritisation and assignment of initial programme theories for testing.

3 | Findings

We identified 96 initial CMO configurations (CMOcs) from the literature, providing a broad evidence base. We then consolidated this evidence into nine candidate programme theories that address diverse international staffing logics. More details on the substantive findings can be found elsewhere (Tate et al. Forthcoming). Briefly, the nine IPTs were as follows:

- IPT 1: Fixed mandated minimum nurse-to-patient ratio. Proposes that legislated ratios with a higher proportion of qualified nurses trigger enhanced clinical competence and rapid responses, leading to improved patient safety and staff retention.
- IPT 2: Sufficient RNs and complement of support staff. Suggests that maintaining adequate Registered Nurse (RN) levels alongside trained assistive personnel reduces non-nursing workloads, allowing RNs to focus on complex clinical coordination.
- IPT 3: Professional nurse substitution. Examines the impact of replacing professional nurses with unregulated

personnel, proposing that remaining RNs feel overburdened as a 'catch-all' for monitoring errors, leading to missed care.

- IPT 4: Temporary nurses unfamiliar with unit culture. Posits that reliance on transient staff unfamiliar with local practices threatens care continuity and demotivates permanent personnel.
- IPT 5: Temporary vs. permanent nursing staff in mental health. Suggests that the use of temporary staff in mental health prevents the formation of therapeutic trust, which is essential for patient safety.
- IPT 6: Classification instrument and nurses' judgement. Explores how failure to meet targets set by acuity tools or professional judgement leads to the omission of essential nursing tasks due to perceived time constraints.
- IPT 7: Overall staffing sufficiency. Proposes that a general lack of personnel and resources compromises professional values and impedes structural empowerment, triggering moral distress.
- IPT 8: Staffing sufficiency and practice norms. Examines how rigid professional boundaries in understaffed environments may lead to essential nursing tasks being neglected by other clinical team members.
- IPT 9: Various skill mixes in long-term care facilities. Proposes that positive resident outcomes in long-term care are driven primarily by the quality of inter-professional relationships rather than a fixed skill mix.

These theories were refined through focus groups with 43 knowledge users who provided input to augment the literature-based findings and prioritize IPTs for each country team. The substantive focus of this refinement process resulted in the prioritization of four key IPTs for global testing: (1) fixed mandated minimum nurse-to-patient ratios, (2) RN-led skill mix with support staff, (3) professional nurse substitution and (4) the frequent use of temporary staff. The findings emphasized conceptual differences within and across country teams and IPTs; for instance, the use of temporary staff was prioritized for theory testing in the Netherlands, but theorized as a 'better than nothing' model essential for survival during the COVID-19 pandemic in Canada, despite its negative impact on teamwork, burnout and care continuity. By presenting these specific theoretical focuses and the rival mechanisms identified by knowledge users, the research provides a transparent audit trail from a large pool of raw data to a focused set of clinically relevant theories. We highlight the process involved for programme theory prioritization in Figure S1.

Our international initial programme theory development approach resulted in 4 initial programme theories for further testing. We agreed that 4 initial programme theories prioritized by participants across all four countries created a meaningful narrative and raised global issues around nurse staffing with sufficient overlap across contexts. We discussed potential options of working on the initial programme theories as a collaborative group, of each team leading on initial programme theory for just their context, or for each country team leading on an initial programme theory but incorporating other country perspectives. We chose the latter option to leverage a unique opportunity of working in a multinational team. This resulted in the final selection of initial programme theories for theory testing in later study phases, which included nursing staffing models characterized by (1) skill mix using a fixed mandated minimum nurse-to-patient ratio (led by the Australian team); (2) skill mix with sufficient RNs and complement of support staff (led by the Canadian team); (3) skill mix through professional nurse substitution to influence quality of care/safety and staff outcomes (led by the British team); and (4) staffing model that focuses on frequent use of temporary nurses who are unfamiliar with a specific unit culture (led by the Dutch team) (see Table 1).

3.1 | Key Learnings From a Multi-National IPT Development Approach

A multi-national approach allowed us to better explore and understand the impact of context in our initial programme theory development. Upon reflecting on our approach, we summarize the key learnings we gained regarding: (1) communicating the value and distinctive features of realist methodology across a mixture of novices, intermediates and methodological experts; (2) achieving a compromise between rigour and relevance in the design of the review itself in working within 'social laboratories' and (3) the logistics of conducting a multinational realist investigation: navigating working remotely, differences in time zones and differences in context and language.

To communicate the value and distinctive features of realist methodology across such a diverse group, we created spaces for mentorship and peer-to-peer learning. We used a combination of learning in advance of doing (through key readings and methodology Q&A sessions), learning by doing (within the individual teams) and learning after doing (in reflective sessions convened across the four country teams). During initial review processes, we intentionally paired senior methodologists and researchers with trainees. We held several multi-national meetings before and during each realist review step to ensure a shared understanding of our aims, philosophical underpinnings and procedures. All research team members were invited to participate in team meetings to encourage partaking in collegial exchanges regardless of level of expertise. This was critical as many research team members were experienced in other review types and had to unlearn as much as learn in order to operationalize a realist approach. Although not explicitly structured this way, our learning was supported through several options for communication; namely group meetings and shared emails threads as well as one-on-one meetings or private email correspondence with team members coordinating the work. Having methodologists and those experienced in realist reviews on each country team resulted in the ability to address team members' questions both within teams and across teams, with everyone benefiting from rich discussions in large team meetings that reconciled or accommodated for differences in training, language and experiences.

Our second lesson relates to reconciling rigour and relevance in the design of our multi-national realist review in working within our local contexts and teams, or 'social laboratories', that cannot and should not be completely controlled and yet which seek to optimize comparison across country contexts. In our regular

TABLE 1 | Prioritized initial programme theory (IPT) summary.

IPT title	Summarized IPT	Complete IPT
IPT 1: Staffing model that focuses on skill mix using a fixed minimum nurse-to-patient ratio	If the nurse staffing model uses fixed minimum nurse-to-patient ratios, with higher ratios of (qualified) nurses to nursing assistants and support workers, as mandated by legislation [context], then this leads to enhanced nurse competence and rapid and appropriate responses [mechanisms] improving the work environment, patient outcomes, and nurse recruitment and retention [outcomes]. (CMOcs 1, 4, 9, 20)	If the nurse staffing model uses fixed minimum nurse-to-patient ratios, with higher ratios of RNs or baccalaureate prepared nurses (versus nursing assistants and support workers), that is mandated by legislation [context] then nurse competence will be enhanced, specifically, there will be early detection of patients at risk of deterioration and quicker initiation of appropriate response[mechanisms] and nurses will perceive there to be more favourable work conditions [mechanism] leading to lower rates of adverse events, and improved patient outcomes[outcomes] and improved nurse outcomes, including improved recruitment and retention of nurses [outcomes]. (CMOcs 1, 4, 9, 20); sources Griffiths et al. (2016); Subirana et al. (2014)
IPT 2: Staffing model that focuses on skill mix with sufficient RNs and complement of support staff	If the nurse staffing model has sufficient RN staffing who have the appropriate skillset, experience and knowledge (maximizing the efficiency of team in a fiscally responsible way; safe nurse to patient ratio dependent on the acuity of the patients) with a complement of assistive nursing personnel who also have the appropriate training or skillset for the work place [context], then RNs will feel more supported in greater responsibilities, confident in their work, spend less time on non-nursing tasks, working relations among the staff are emphasized, and perceptions of sufficient staffing may improve to reduce stress among staff [mechanisms], which will improve staff outcomes, quality of care, and lower hospital costs [outcomes]. (CMOcs 7, 10, 18). *Rival theory:* *If the nurse staffing model has sufficient RN staffing who have the appropriate skillset, experience and knowledge (maximizing the efficiency of team in a fiscally responsible way; safe nurse to patient ratio dependent on the acuity of the patients) with a complement of assistive nursing personnel who also have the appropriate training or skillset for the work place [context], then RNs will feel like they are not engaging in relational practice, will worry about appropriate task sharing with LPNs, and will feel that the support by assistive personnel is not adequate or appropriate for acutely unwell patients (e.g., neuro patients) [mechanisms], which will lead to decreased job satisfaction, lack of role clarity, and duplication of tasks respectively [outcomes] (CMOcs 7, 10, 18)*	If the nurse staffing model has sufficient RN staffing who have the appropriate skillset, experience and knowledge (maximizing the efficiency of team in a fiscally responsible way; safe nurse to patient ratio dependent on the acuity of the patients) on units with a complement of assistive nursing personnel who also have the appropriate training or skillset for the work place where it is understood that the RN is responsible for certain patient responsibilities [specified as indirect care coordination responsibilities such as delegation of work to unregulated workers and ensuring clarity between the roles of all of the patient care providers] [context], then RNs feel more supported to have greater responsibilities, confident in their work, and spend less time on non-nursing tasks, working relations among the staff are emphasized, and perceptions of sufficient staffing may improve to reduce stress among staff [mechanisms], leading to RNs performing well in areas they were exclusively trained for, improved healthcare provider relationships, job satisfaction, higher composite quality index score, streamlined nursing care processes, and lower turnover with subsequent lower hospital costs [outcomes] (CMOcs 7, 10, 18); Sources: Hall (1998); O'Brien-Pallas et al. (2011); Spilsbury et al. (2011)

(Continues)

TABLE 1 | (Continued)

IPT title	Summarized IPT	Complete IPT
IPT 3: Staffing model that focuses on skill mix through professional nurse substitution to influence quality of care/safety and staff outcomes	If the nurse staffing model increases the proportion of assistive nursing personnel in acute care, especially when patients are older and have more comorbidities [context], then nurses with professional knowledge will use that professional knowledge to comprehensively understand the clinical presentation of the patient, but will feel overburdened and pressured to be the 'catch-all' (the one monitoring and catching any errors) [mechanism], leading to lower quality of care, turnover, poor staff outcomes, poor patient outcomes (i.e., missed or poor care, errors made in care), and increased cost [outcomes] (CMOcs 3, 12, 15)	If the nurse staffing model increases the proportion of assistive nursing personnel without professional nurse qualifications (versus professional nurses) in response to low nurse staffing levels in acute care, especially when patients are older and have more comorbidities [context] then less knowledge stock (professional knowledge) will be employed [mechanism] and professional nurses will use that professional knowledge to comprehensively understand the clinical presentation of the patient, but will feel overburdened and pressured to be the 'catch-all' (the one monitoring and catching any errors) [mechanism], leading to preventable deaths and trade-offs (using less educated staff or changing nurse patient ratios) that result in lower quality of care, poor patient outcomes (i.e., missed or poor care, errors made in care), higher turnover, poor retention, an increase in the use of overtime, on-call pay, and sick leave for the nonprofessional staff [outcomes] leading to a higher cost of nursing care [outcome] (CMOc 3, 12, 15); Sources: Hall et al. (2004); Leary and Punshon (2019)
IPT 4: Staffing model that frequently uses temporary nurses who are unfamiliar with a specific unit culture	If the nurse staffing model frequently uses non-full time and temporary nurses who are unfamiliar with a unit and who may not have the skills required for the unit[context], then this threatens care continuity and permanent RN staff will feel overburdened and demotivated, while temporary/part-time staff will not know the patient/client or their family [mechanisms], leading to poor teamwork, burnout, more adverse events, longer hospital stays and increase costs [outcomes] (CMOcs 11, 13, 14). **better than nothing model in COVID-19 pandemic*	If the nurse staffing model frequently uses non-full time and temporary nurses who are unfamiliar with a specific unit culture and practices (versus an adequate complement of permanent nurses) and who may not have the skills required for the unit [context] then this threatens care continuity [mechanism], permanent RN staff will feel overburdened, experience pressure and demotivation, while temporary/part-time staff will not know the patient/client or their family [mechanism], leading to poor teamwork, burnout, turnover, longer length of hospital stays, higher rates of adverse events (including death, pneumonia, cardiac arrest) and increases in health care costs [outcomes]. (CMOcs 11, 13, 14); Sources: Baker et al. (2019); O'Brien-Pallas et al. (2011)

meetings, we discussed which aspects of the review should be consistent across all countries and opportunities in which we should capitalize on our large, multinational team to develop specific and contextualized understandings of our initial programme theories. We had consistent and clear inclusion and exclusion criteria for our review and met regularly to ensure a shared understanding of nurse staffing models, mid-range theory and theoretical statements or CMOcs to aid in our screening processes. We engaged in shared learning about what nurse staffing models and nursing positions or titles meant across different contexts. During this time, we shared resources, emails on and discussed in online meetings, the differences across contexts while maintaining a broad, but consistent focus for the inclusion and extraction of articles for CMOcs that would later be synthesized into initial programme theories. We worked across country teams to conduct preliminary synthesis and discuss a large amount of CMOcs, but then intentionally assigned subsequent stakeholder engagement work for IPT prioritization to be focused within each country team (while supporting open communication and touchpoints about process and findings). Notably, we ensured flexibility regarding recruitment for stakeholder engagement by having the most appropriate individuals from each context speak to how nurse staffing models are thought to operate and produce outcomes among countries. Therefore, each country team selected the type of knowledge user most relevant to their context to maintain the integrity of the intent of the review, rather than ensure everything was standardized. We prioritized initial programme theories within each country via stakeholder engagement but came together as a multi-national team to confer about the assignment of initial programme theories across countries for subsequent research steps. We maintained rigour by ensuring certain procedures were consistent across country teams (e.g., initial prioritization procedures, focus group guides) and we had a shared understanding of process and project goals, while making sur this consistency did not interfere with our ability to develop rich, *contextualized* descriptions of how nurse staffing models are thought to work. This also allowed us to identify questions for stakeholder clarification across different contexts as well as commonalities in initial programme theories for which findings may be transferable.

Challenges navigating the logistics of a large, multinational team resulted in key insights into what worked and what did not. For example, in the process of CMOc consolidation and initial prioritization, a strategy for reviewing and discussing the extracted CMOcs was determined in advance and was led by the team with the nominated principal applicant (the Canadian team). Often, with clear instructions and endpoints, we believe this helped the research process. A team member from the Canadian team consistently chaired the meetings and consolidated work across countries. When the entire team (all four countries) was present, the process could be immediately discussed within the group. When the large group was divided into subgroups (e.g., the initial prioritization working meeting with the research team), the process in both groups sometimes deviated based on the nature of the data being worked through or group dynamics, potentially impacting the intended outputs of our work. Within this four-country research group, a substantial amount of new knowledge was generated (at different moments and in different team compositions) and this knowledge was continually shared in the research group. It was important to constantly and explicitly capture this knowledge by developing glossaries and utilizing tables, meeting minutes and emails. However, maintaining everyone was 'up to date' and 'on the same page' was challenging. This was particularly demanding in a realist review where certain aspects of the upcoming steps are not always predetermined (nor should be). In these cases, it was crucial to have an individual or a subgroup capable of maintaining an overview and keeping everyone engaged; having an experienced methodologist (and in our study, multiple methodologists) was essential. Of note, we could only partially address logistic challenges around time zones by holding meetings to ensure that, for at least parts of the year, one team did not have only late-night meetings. Differences in costs considered eligible by the sponsor and across different institutions also impacted how individuals could be compensated fairly and therefore involved in different aspects of the work.

4 | Discussion

Through our work, we gained key insights from our initial programme theory development, prioritization and assignment. Capitalizing on differences within and across multi-national team contexts is essential for any form of comparative research addressing complex health and social issues (Kosmützky 2018). During this phase of our realist review, we identified aspects of our work in which we required a consistent approach across teams as well as aspects of the work where we not only could but should leave room for team and contextual differences to be optimized.

Our large team's capacity enabled a more systemic and comprehensive review of literature, expanding the pool of candidate theories for our multi-national project. Importantly, the 'social laboratories' or country teams themselves (i.e., the existence of these four country teams and their composition) influenced the identification of contexts, mechanisms and outcomes related to various nurse staffing models, shaping our interpretation of findings and assignment of programme theories. Interactions across the four sub-groups influenced the way our research approach was implemented, reflecting notions by other authors on the hermeneutic nature of responsible research innovation, especially when focusing on social change (Timmermans et al. 2020). In the spirit of our realist approach and to capitalize on contextual differences across teams, we worked together to identify the demi-regularities of our process that worked to produce the outcomes and outputs of interest in this study (Jagosh et al. 2012). Some of this was intentional, such as, agreeing on differences in participant recruitment across contexts. Other differences in our country team approaches were products of differences in team composition, expertise and contextual knowledge that did not emerge as relevant until within country work occurred, such as when we split into two subgroups for the initial prioritization of CMOcs for preliminary synthesis. In this vein, we recommend embracing a rhizomatic way of working (Fox 2023), but stress the need for open communication and consistent check-ins for progress updates that also serve as venues for expectation and goal-setting (Crites et al. 2020). Echoing Flynn et al. (2020), we recommend narrowing the scope of your

work to allow for in-depth exploration of context, even with a large, multi-national team (Flynn et al. 2020).

Our international work prompted us to reflect on and re-examine the terms, frameworks and assumptions about our health systems and workforce. Our cross-country collaboration enabled a recognition of the familiar and allowed us to make explicit those latent characteristics of global research in this field. We could not rely on nursing titles to be consistent reflections of the training, education, scope of practice or roles of nurses in each country. Moreover, a challenge we continually faced throughout this review was the varying descriptions, or lack of clear and specific descriptions of a nurse staffing model. This is likely due to the context-dependent nature of a nurse staffing model across care settings, patient needs and historical/political environments (Matthias 2015; Schalkwijk et al. 2024; Tate et al. 2024). These challenges befit a realist approach but still require a firm setting of parameters and procedures to ensure the work is feasibly completed. We recommend ensuring constant and thorough reflection on the clarity of the model, program or intervention of interest (Jagosh et al. 2014). Where we could identify clear, consistent definitions in the literature, we had to ensure we made note to ask our participants about the meaning and operation of certain titles or concepts in each of their contexts (e.g., having sufficient nursing staffing was mentioned often in the literature but not defined). We codified knowledge gained through iterative revisions, additions and clarifications to study materials such as inclusion criteria, tables of definitions, meeting minutes, process instructions and technical reports (Crites et al. 2020).

The international composition of our realist review team was a valuable asset, reflecting the complexity of the study and individual IPTs. The diverse perspectives, experiences and methodological expertise contributed to a rich working environment. Other realist syntheses have, of course, benefited from this international abundance on many occasions, but it is worth noting how our own collaboration might help evolve the field. It should first be viewed in the context of a global virtual team (GVT). The GVT has become established as an efficient and productive mode of working since the COVID pandemic. On one level we can state that the establishment and maintenance of a GVT for the purpose of realist synthesis was successful. Although the work was divided into a small number (four) of individual reviews, the connections and overlap, shared practice and decision making were such that it constituted a singular venture, inhabiting a common set of objectives. The practices of our GVT can be applied to future international realist endeavours, including regular communication, common processes, conventional approaches to documentation and a shared commitment to asynchronous work and milestones. The shared identity and mutual support within our GVT facilitated the completion of the work and fostered a sense of team identity, crucial for realist synthesis. This GVT identity was forged in the process of the work but became particularly apparent when discussion about our common observations on meta-narrative began to emerge.

We recommend being intentional in creating spaces for mentorship and peer learning for team building and capacity building. This is important for generating knowledge and skill in realist methodologies and shared cultural competencies. We recommend acknowledging and addressing the various diversities within and across teams, such as educational background, training level and institutional norms, which can impact collaborative work (Kosmützky 2018). Team composition across and within country teams was important, and it was especially key that the individuals enacted the roles they set out to do in the initial grant application for this work. Although much of the 'hands-on' work was completed by trainees, junior faculty and paid staff, consistent engagement, rather than just consultation, by senior researchers and methodologists throughout the research process was key. For each step in the research process, we employed an egalitarian approach for assigning workloads for review processes and prioritization procedures; enhanced the quality of the work produced; and made for a more enriching learning and work environment (Crites et al. 2020; Kosmützky 2018). This norm within our working group was largely driven by the senior researchers and methodologists willing to engage and build capacity within our team, reflecting the leadership and stewardship espoused as critical for collaborative learning environments online (Crites et al. 2020). One final feature of this capacity building through the creation of space and opportunity for discussion exists in relation to our growing awareness of other countries' contexts. This was associated not just with healthcare as the focus of our energy but also academic environments within which we work, and included operational, cultural and social circumstances.

Embracing cultural communication and team building was needed to address the complexities of international comparative and collaborative research (Kosmützky 2018). Despite some prior collaboration, many team members had not met before this project. Initially, there was uncertainty in navigating social dynamics within the team. Team meetings were followed by email communication to clarify norms, professional relationships and roles of team members, realist approaches and terminology to the team coordinator. For those not wanting to take up limited meeting time, those unsure of their questions or those who had insights after reflection upon group discussions, encouraging open communication allowed for questions to be asked or opinions to be voiced that could be clarified individually or brought up to the larger group later for shared learning, as appropriate. Trainee meetings allowed for sense- making and the identification of knowledge gaps that could be brought forward to senior researchers in solidarity. The positive response to these questions and concerns cultivated a virtual environment of curious inquiry. It is worth noting that, when considering the initial meetings, team members used silence to protect themselves (i.e., from potential embarrassment) which should be considered a sign of being considerate of and cultivating internal relationships among the team. Indeed, the role of silence, or better, what is said and what is not said adds to the discussion in interdisciplinary research teams through the subtext of consent or resistance it can convey (Verouden et al. 2016). Especially among international research teams, we recommend being cognizant of who is silent and when as indicators of internal relationships, team function and whether team members or recognizing and being responsive to differences in culturally bound knowledge (Kosmützky 2018).

5 | Conclusions and Implications

This article reports on the experience of working in a multi-national team to conduct a realist review on which nurse staffing models

work, for whom and in what contexts. Our approach allowed for the resources, expertise and contextualized perspectives necessary to explore and develop multiple initial programme theories around nurse staffing. Having strong engagement from senior researchers and experts who have contextually specific knowledge and methodological expertise that goes beyond the knowledge of procedures and promotes the principles of realist methods was key. Moreover, the process used to develop programme theories demonstrates a replicable approach for conducting realist reviews in complex, multi-national settings, highlighting the value of iterative learning, diversity of perspectives and collaborative synthesis. We recommend establishing and agreeing upon certain routines, procedures and check-ins as a diverse, multi-national team, but also emphasize embracing a rhizomatic way of working that optimizes different strengths and culturally bound knowledge to enhance rigour in realist approaches.

Author Contributions

K.T., A.B., G.H., T.R., I.W., P.L., C.O., R.F., T.P., M.H., T.S., G.G.C. and members of the ReSoNANCE (Realist Synthesis of Nursing in Australia Netherlands Canada and England) group contributed to the conceptualization, writing, review and/or editing of the manuscript.

Acknowledgements

The authors acknowledge the active partnership and full contribution of the research team of decision-makers and researchers as listed below: Principal Applicants & Project Advisory Team Greta Cummings, Nominated Principal Applicant, University of Alberta. Christy Raymond, Principal Applicant, MacEwan University. Alison Kitson, Principal Applicant, Flinders University. Gillian Harvey, Principal Applicant, Flinders University. Pieterbas Lalleman, Principal Applicant, Fontys University. Lisette Schoonhoven, Principal Applicant, University of Applied Science Utrecht. Andrew Booth, Principal Applicant, The University of Sheffield. Tony Ryan, Principal Applicant, The University of Sheffield. Kaitlyn Tate, Principal Applicant, University of Alberta. Independent Researcher Co-Applicants Carole Estabrooks Co-Applicant, University of Alberta. Malcolm Doupe Co-Applicant, University of Manitoba. Matthias Hoben Co-Applicant, York University. Rachel Flynn Co-Applicant, University College Cork. Steve Robertson Co-Applicant, The University of Sheffield. Angela Tod Co-Applicant, The University of Sheffield. Tim Schultz Co-Applicant, Flinders University. Catharina J van Oostveen Co-Applicant, Erasmus University. Knowledge User Co-Applicants Alison Hutchinson Co-Applicant, Deakin University. Sean Chilton Co-Applicant, Alberta Health Services. Deb Gordon Co-Applicant, Alberta Health Services. Mike Villeneuve Co-Applicant, Canadian Nurses Association (Ottawa)/Association des infirmières et infirmiers du Canada (Ottawa). Wilma Jackson Co-Applicant, Elisabeth-Tweestedenziekenhuis Hospital. Dewi Stalpers Co-Applicant, Dutch Hospital Association and St. Antonius Hospital. Sam Debbage Co-Applicant, Doncaster and Bassetlaw NHS Teaching Hospitals Foundation Trust. Trainees & Staff Inge Wolbers, Doctoral Trainee, University of Applied Science Utrecht. Tatiana Penconek, Doctoral Trainee, University of Alberta. Shovana Shrestha, Doctoral Trainee, University of Alberta. Sadie Deschenes, PhD, University of Alberta. Patrick Chiu, PhD, University of Alberta. Additional Members Candice Oster, Flinders University. Lemma Bulto, Flinders University. Katherine Jones, The University of Sheffield, ReSoNANCE (Realist Synthesis of Nursing in Australia Netherlands Canada and England).

Funding

This study was funded by a grant from the Canadian Institutes of Health Research (#169127) to Greta G. Cummings.

Ethics Statement

This study was approved by the University of Alberta Research Ethics Committee (approval no. Pro00100425) on May 20, 2024.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data produced in the present study are available upon reasonable request to the authors.

References

Aiken, L. H., D. M. Sloane, L. Bruyneel, et al. 2014. "Nursing Staffing and Education and Hospital Mortality in Nine European Countries: A Retrospective Observational Study." *Lancet* 383, no. 9931: 1824–1830. https://doi.org/10.1016/S0140-6736(13)62631-8.

Astbury, B., and F. L. Leeuw. 2010. "Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation." *American Journal of Evaluation* 31, no. 3: 363–381.

Baker, J. A., K. Canvin, and K. Berzins. 2019. "The Relationship Between Workforce Characteristics and Perception of Quality of Care in Mental Health: A Qualitative Study." *International Journal of Nursing Studies* 100: 103412. https://doi.org/10.1016/j.ijnurstu.2019.103412.

Boniol, M., T. Kunjumen, T. S. Nair, A. Siyam, J. Campbell, and K. Diallo. 2022. "The Global Health Workforce Stock and Distribution in 2020 and 2030: A Threat to Equity and 'Universal' Health Coverage?" *BMJ Global Health* 7: e009316.

Bridges, J., P. Griffiths, E. Oliver, and R. M. Pickering. 2019. "Hospital Nurse Staffing and Staff–Patient Interactions: An Observational Study." *BMJ Quality and Safety* 28, no. 9: 706–713. https://doi.org/10.1136/bmjqs-2018-008948.

Butler, M., T. J. Schultz, P. Halligan, et al. 2019. "Hospital Nurse-Staffing Models and Patient- and Staff-Related Outcomes." *Cochrane Database of Systematic Reviews* 2019, no. 4: CD007019. https://doi.org/10.1002/14651858.cd007019.pub3.

Crites, G. E., A. Berry, E. Hall, D. Kay, M. K. Khalil, and L. Hurtubise. 2020. "Applying Multiple Frameworks to Establish Effective Virtual Collaborative Teams in Academia: A Review and Recommendations." *Medical Education Online* 25, no. 1: 1742968. https://doi.org/10.1080/10872981.2020.1742968.

Dada, S., S. Dalkin, B. Gilmore, R. Hunter, and F. C. Mukumbang. 2023. "Applying and Reporting Relevance, Richness and Rigour in Realist Evidence Appraisals: Advancing Key Concepts in Realist Review." *Research Synthesis Methods* 14, no. 3: 504–514.

Dall'Ora, C., C. Saville, B. Rubbo, L. Y. Turner, J. Jones, and P. Griffiths. 2022. "Nurse Staffing Levels and Patient Outcomes: A Systematic Review of Longitudinal Studies." *International Journal of Nursing Studies* 134: 104311. https://doi.org/10.1016/j.ijnurstu.2022.104311.

Flynn, R., K. Schick-Makaroff, A. Levay, and J. Greenhalgh. 2020. "Developing an Initial Program Theory to Explain How Patient-Reported Outcomes Are Used in Health Care Settings: Methodological Process and Lessons Learned." *International Journal of Qualitative Methods* 19: 160940692091629. https://doi.org/10.1177/1609406920916299.

Fox, N. J. 2023. "Rhizomatic Review: A Materialist Minor Science Approach to Research Evaluation." *Qualitative Research* 24, no. 5: 1101–1120. https://doi.org/10.1177/14687941231216632.

Griffiths, P., J. Ball, J. Drennan, et al. 2016. "Nurse Staffing and Patient Outcomes: Strengths and Limitations of the Evidence to Inform Policy and Practice. A Review and Discussion Paper Based on Evidence Reviewed for the National Institute for Health and Care

Excellence Safe Staffing Guideline Development." *International Journal of Nursing Studies* 63: 213–225. https://doi.org/10.1016/j.ijnurstu.2016.03.012.

Hall, L. M. 1998. "Policy Implications When Changing Staff Mix." *Nursing Economic* 16, no. 6: s291–s312.

Hall, L. M., D. Doran, and G. H. Pink. 2004. "Nurse Staffing Models, Nursing Hours, and Patient Safety Outcomes." *Journal of Nursing Administration* 34, no. 1: 41–45. https://doi.org/10.1097/00005110-200401000-00009.

Imam, A., S. Obiesie, J. Aluvaala, J. M. Maina, D. Gathara, and M. English. 2022. "Identifying Gaps in Global Evidence for Nurse Staffing and Patient Care Outcomes Research in Low/Middle-Income Countries: An Umbrella Review." *BMJ Open* 12, no. 10: e064050. https://doi.org/10.1136/bmjopen-2022-064050.

International Council of Nurses. 2023. "International Supply of Nurses." https://icn.ch/sites/default/files/2023-04/COVID19_internationalsupplyofnurses_Report_FINAL.pdf.

Jagosh, J., A. C. Macaulay, P. Pluye, et al. 2012. "Uncovering the Benefits of Participatory Research: Implications of a Realist Review for Health Research and Practice." *Milbank Quarterly* 90, no. 2: 311–346. https://doi.org/10.1111/j.1468-0009.2012.00665.x.

Jagosh, J., P. Pluye, A. C. Macaulay, et al. 2011. "Assessing the Outcomes of Participatory Research: Protocol for Identifying, Selecting, Appraising and Synthesizing the Literature for Realist Review." *Implementation Science* 6: 24. https://doi.org/10.1186/1748-5908-6-24.

Jagosh, J., P. Pluye, G. Wong, et al. 2014. "Critical Reflections on Realist Review: Insights From Customizing the Methodology to the Needs of Participatory Research Assessment." *Research Synthesis Methods* 5, no. 2: 131–141. https://doi.org/10.1002/jrsm.1099.

Kosmützky, A. 2018. "A Two-Sided Medal: On the Complexity of International Comparative and Collaborative Team Research." *Higher Education Quarterly* 72: 314–331. https://doi.org/10.1111/hequ.12156.

Leary, A., and G. Punshon. 2019. "Determining Acute Nurse Staffing: A Hermeneutic Review of an Evolving Science." *BMJ Open* 9, no. 3: e025654. https://doi.org/10.1136/bmjopen-2018-025654.

Long, C. J., N. M. Sarkies, E. F. Auton, et al. 2022. "Conceptualising Contexts, Mechanisms and Outcomes for Implementing Large-Scale, Multisite Hospital Improvement Initiatives: A Realist Synthesis." *BMJ Open* 12: e058158. https://doi.org/10.1136/bmjopen-2021-058158.

Marschalek, I., V. Blok, M. Bernstein, et al. 2022. "The Social Lab as a Method for Experimental Engagement in Participatory Research." *Journal of Responsible Innovation* 9, no. 3: 419–442. https://doi.org/10.1080/23299460.2022.2119003.

Matthias, A. D. 2015. "Making the Case for Differentiation of Registered Nurse Practice: Historical Perspectives Meet Contemporary Efforts." *Journal of Nursing Education and Practice* 5: 108.

O'Brien-Pallas, L., R. M. Meyer, L. J. Hayes, and S. Wang. 2011. "The Patient Care Delivery Model—An Open System Framework: Conceptualisation, Literature Review and Analytical Strategy." *Journal of Clinical Nursing* 20, no. 11–12: 1640–1650. https://doi.org/10.1111/j.1365-2702.2010.03391.x.

Pawson, R., T. Greenhalgh, G. Harvey, and K. Walshe. 2005. "Realist Review: A New Method of Systematic Review Designed for Complex Policy Interventions." *Journal of Health Services Research & Policy* 10: 22–34.

Project, R. I. 2017. "Realist Interview and Realist Evaluation Interviewing-A 'Starter Set' of Questions."

Roush, K. 2020. "The Impact of COVID-19 on the Global Nursing Workforce." *American Journal of Nursing* 120, no. 11: 14–15. https://doi.org/10.1097/01.NAJ.0000721876.75249.68.

Schalkwijk, H., M. Felder, P. Lalleman, M. S. Parry, L. Schoonhoven, and I. Wallenburg. 2024. "Five Pathways Into One Profession: Fifty Years of Debate on Differentiated Nursing Practice." *Nursing Inquiry* 31: e12631. https://doi.org/10.1111/nin.12631.

Sherenian, M., J. Profit, B. Schmidt, et al. 2013. "Nurse-To-Patient Ratios and Neonatal Outcomes: A Brief Systematic Review." *Neonatology* 104, no. 3: 179–183. https://doi.org/10.1159/000353458.

Smeets, R. G. M., D. F. L. Hertroijs, F. C. Mukumbang, M. E. A. L. Kroese, D. Ruwaard, and A. M. J. Elissen. 2022. "First Things First: How to Elicit the Initial Program Thoery for a Relaist Evaluation of Complex Integrated Care Programs." *Milbank Quarterly* 100: 151–189. https://doi.org/10.1111/1468-0009.12543.

Spilsbury, K., C. Hewitt, L. Stirk, and C. Bowman. 2011. "The Relationship Between Nurse Staffing and Quality of Care in Nursing Homes: A Systematic Review." *International Journal of Nursing Studies* 48, no. 6: 732–750. https://doi.org/10.1016/j.ijnurstu.2011.02.014.

Subirana, M., A. Long, J. Greenhalgh, and J. Firth. 2014. "A Realist Logic Model of the Links Between Nurse Staffing and the Outcomes of Nursing." *Journal of Research in Nursing* 19, no. 1: 8–23. https://doi.org/10.1177/1744987113481782.

Tamata, A. T., and M. Mohammadnezhad. 2023. "A Systematic Review Study on the Factors Affecting Shortage of Nursing Workforce in the Hospitals." *Nursing Open* 10: 1247–1257. https://doi.org/10.1002/nop2.1434.

Tate, K., A. Booth, G. Harvey, et al. Forthcoming. "Developing Initial Programme Theories in a 4-Country Realist Review of Contextually Appropriate Nurse Staffing Models: Findings From the Canadian Context." *International Journal of Nursing Studies*.

Tate, K., T. Penconek, A. Booth, et al. 2024. "Contextually Appropriate Nurse Staffing Models: A Realist Review Protocol." *BMJ Open* 14, no. 5: e082883. https://doi.org/10.1136/bmjopen-2023-082883.

Timmermans, J., V. Blok, R. Braun, R. Wesselink, and R. Ø. Nielsen. 2020. "Social Labs as an Inclusive Methodology to Implement and Study Social Change: The Case of Responsible Research and Innovation." *Journal of Responsible Innovation* 7, no. 3: 410–426. https://doi.org/10.1080/23299460.2020.1787751.

van Oostveen, C. J., E. Mathijssen, and H. Vermeulen. 2015. "Nurse Staffing Issues Are Just the Tip of the Iceberg: A Qualitative Study About Nurses' Perceptions of Nurse Staffing." *International Journal of Nursing Studies* 52, no. 8: 1300–1309. https://doi.org/10.1016/j.ijnurstu.2015.04.002.

Verouden, N. W., M. C. van der Sanden, and N. Aarts. 2016. "Silence in Interdisciplinary Research Collaboration: Not Everything Said Is Relevant, Not Everything Relevant Is Said." *Science as Culture* 25, no. 2: 264–288.

Wong, G., T. Greenhalgh, G. Westhorp, J. Buckingham, and R. Pawson. 2013. "RAMESES Publication Standards: Realist Syntheses." *BMC Medicine* 11, no. 21: 1–14.

Wong, G., T. Greenhalgh, G. Westhorp, and R. Pawson. 2012. "Realist Methods in Medical Education Research: What Are They and What Can They Contribute?" *Medical Education* 46: 86–96.

World Health Organization. 2024. "Nursing and Midwifery." https://www.who.int/news-room/fact-sheets/detail/nursing-and-midwifery.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Initial programme theory development & prioritization: a 4-stage process. **File S1:** Focus group interview guide: health decision-makers/expert stakeholders.