BMJ Open

Realist evaluation of UK medical education quality assurance

Paul Crampton [ORCID],[1,2] Leila Mehdizadeh,[2] Michael Page [ORCID],[3] Laura Knight,[2] Ann Griffin[2]

[1]Health Professions Education Unit, Hull York Medical School, York, UK
[2]Research Department for Medical Education, UCL, London, UK
[3]Institute for Health Sciences Education, QMUL, London, UK

Correspondence to
Professor
Ann Griffin; a.griffin@ucl.ac.uk

ABSTRACT

Objectives The aim of the study was to explore what components of the General Medical Council's (GMC) Quality Assurance Framework work, for whom, in what circumstances and how?

Setting UK undergraduate and postgraduate medical education and training.

Participants We conducted interviews with a stratified sample of 36 individuals. This included those who had direct experiences, as well as those with external insights, representing local, national and international organisations within and outside medicine.

Intervention The GMC quality assure education to protect patient and public safety utilising complex intervention components including meeting standards, institutional visits and monitoring performance. However, the context in which these are implemented matters. We undertook an innovative realist evaluation to test an initial programme theory. Data were analysed using framework analysis.

Results Across components of the intervention, we identified key mechanisms, including transparent reporting to promote quality improvement; dialogical feedback; partnership working facilitating interactions between regulators and providers, and role clarity in conducting proportionate interventions appropriate to risk. The GMC's framework was commended for being comprehensive and enabling a broad understanding of an organisation's performance. Unintended consequences included confusion over roles and boundaries in different contexts which often undermined effectiveness.

Conclusions This realist evaluation substantiates the literature and reveals deeper understandings about quality assuring medical education. While standardised approaches are implemented, interventions need to be contextually proportionate. Routine communication is beneficial to verify data, share concerns and check risk; however, ongoing partnership working can foster assurance. The study provides a modified programme theory to explicate how education providers and regulators can work more effectively together to uphold education quality, and ultimately protect public safety. The findings have influenced the GMC's approach to quality assurance which impacts on all medical students and doctors in training.

Strengths and limitations of this study

► The study is underpinned by a sound team-based reflexive analysis and is the first in-depth realist evaluation of quality assurance in the healthcare context.
► The study fills a critical knowledge gap on quality assurance and the findings have influenced the General Medical Council's approach which impacts on all medical students and doctors in training.
► To enhance the transferability of our findings, we collected qualitative data from a large number of UK and international based expert stakeholders within and outside medicine.
► A limitation is that the study is specific in its UK focus and focuses on the General Medical Council.
► Furthermore, the study did not analyse the impact of outcome data (eg, National Training Survey, Care Quality Commission data) to understand links between quality assurance and intervention components.

setting standards, institutional visits and monitoring performance. However, the context in which these components are implemented matters.[1–3] Within undergraduate and postgraduate medical education and training, the taught curriculum integrates with workplace-based experiential learning. Consequently, education environments range from structured classrooms in university contexts to clinical placements within shifting healthcare structures across primary and secondary care. Therefore, the challenge for regulators is to mediate the quality of education and training across these spaces in order to assure the public that education and training is safe, that all medical students are prepared for practice and that all trainee doctors are fit to practise.

In the UK context, the General Medical Council (GMC) regulator work closely with other organisations to secure its standards, using a three-tier model. The GMC (tier 1, quality assurance (QA)), has an overarching responsibility to hold undergraduate and postgraduate training bodies to account for meeting standards. These bodies (tier 2,
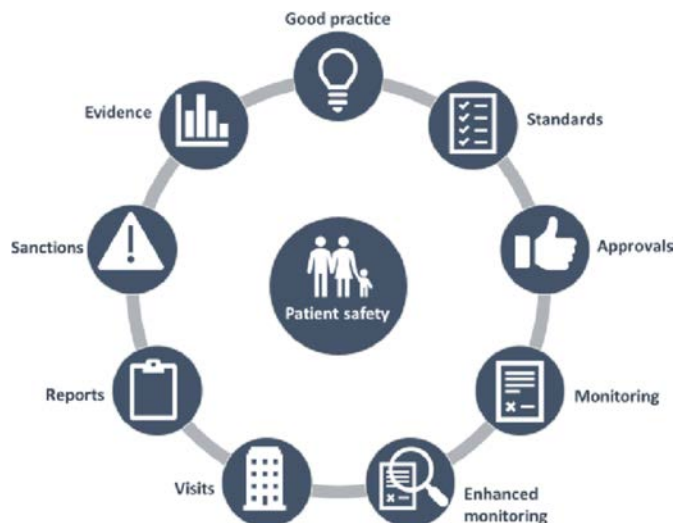
INTRODUCTION
Rationale

Healthcare regulators quality assure education and protect public safety utilising complex intervention components including

BMJ

Figure 1 The intervention: Quality Assurance Framework.

quality management) organise, manage, commission and sometimes deliver medical education. They also manage quality in local education providers (LEPs), where students and trainees are placed, such as trusts, health boards, general practices and other clinical settings. The LEPs (tier 3, quality control) have processes to ensure satisfactory clinical placements, and that their organisation provides an appropriate learning environment. Medical Royal Colleges work with the GMC to ensure their curricula and assessments are fit for purpose, inform specialty and postgraduate programme delivery, and have local systems to support training.

The GMC has a multifaceted intervention to examine the quality of medical education and training provision, known as the Quality Assurance Framework (QAF, figure 1). The intervention includes the following components: setting standards, approving education settings, monitoring activities including self-assessment and enhanced monitoring, visits, sharing evidence with other regulators and identifying good practice.[4] The QAF operates across the three-tier model that is, between the regulator (eg, GMC), organisational provider bodies (eg, medical schools) and local service delivery organisations (eg, hospitals, general practice).

A range of approaches can be implemented to assure education quality, from heavily arbitrated measures to informal uncontrolled processes. Existing education QA research is sparse and tends to come from the field of school-based and higher education.[5 6] While exploring the mechanisms of action of school inspections,[5] a theory stated that regulatory activities associated with improvement include: setting standards, the provision of feedback, employing a system of sanctions and rewards, monitoring schools by the collection of information and public accountability. However, more research is needed to understand within the healthcare context how QA can protect patients. Despite large amounts of resources dedicated to education QA, there remains a lack of clear evidence. QA takes place within varied and complex

social environments. For this reason, the same intervention can impact on individuals, teams and organisations in different ways.[7 8] Although there are intended consequences explicit in the QAF design, the implicit underlying drivers of these are not clear.

Specific aim
The study aim was to explore what components of the GMC's QAF work, for whom, in what circumstances and how?

METHODOLOGY
Conceptual framework: realist evaluation
We chose realist methodology because of a focus on four theoretically constructed and inter-related questions: what works, for whom, in what circumstances and how?[9–11] This results in generative causation, about how QAF components operate, offering an assessment of *whether* they work, as well as *why*. In the results, we explore the complex configuration links between contexts (where, when and with whom the activity takes place), components (different activities applied to assure quality), mechanisms (underlying processes for why the activity is/is not effective) and outcomes (intended and unintended consequences).[12] Our methods follow the Realist and Meta-narrative Evidence Synthesis: Evolving Standards 2[12] reporting guidelines (for full report, see Griffin *et al*).[13]

Initial programme theory
We developed an initial programme theory (figure 2) based on existing literature, the GMC's approach to QA and research team insight (see the Reflexivity section). We positioned the QAF as consisting of various components which we then explored to answer our study aim. Each component triggers multiple responses when applied in certain contexts with underpinning resources. Our programme theory postulated that within undergraduate, postgraduate and LEP contexts, the QAF led to improved quality and protected the public through exploiting regulatory influence, guidance and supporting organisations, leading to compliance, resistance, relationships and empowerment.

Sampling and recruitment
Stratified purposive sampling was used to test our theory with stakeholders. We targeted those who were familiar with the QAF, labelled in the study as QA partners (QAPs); as well as those with outsider perceptions of how the framework is positioned in society (eg, other regulators) and broader regulation contexts (eg, education). These spanned organisations both inside and outside healthcare, including internationally; collectively labelled as non-QA partners (non-QAPs). See table 1 for the list of QAPs.

Our realist position acknowledged that stakeholders each had partial knowledge of the intervention, therefore to fully explore research questions we included
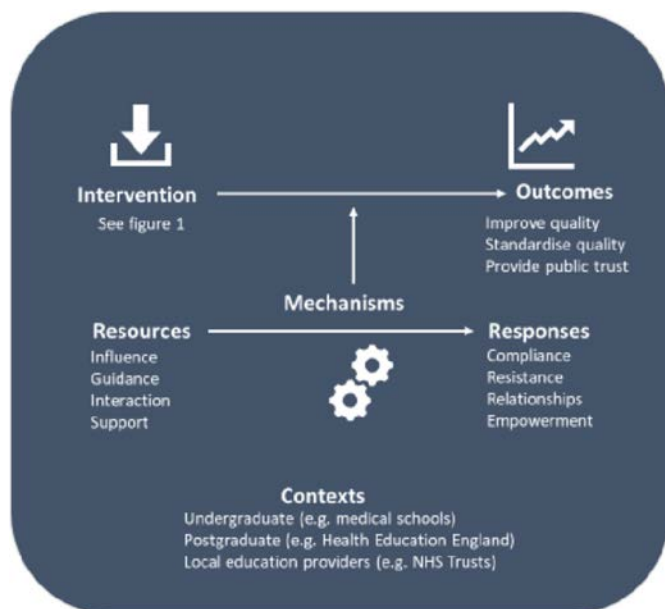
Figure 2 Initial programme theory. NHS, National Health Service.

policy-makers, implementers and recipients.[14] Following email invitations, non-responders received two reminders. Aligned to realist evaluation,[15] participants were given the opportunity to review project materials prior to participation via a 15 min informational video.

Data collection

To test our programme theory, we undertook semi-structured interviews to explore underlying processes triggered by QA components (online supplementary file 1). We tested a number of candidate theories to explore the underlying ways in which the intervention was intended to be successful. For example, to test theories on the impact of generalisability and utility,[16] we asked: "Some hold the view that organisational self-assessment is not a reliable process, what do you think?", and "What would happen if the medical regulator (GMC) did no organisational visiting?" We designed questions with different foci appropriate to participants,[15] for QAPs and non-QAPs. Additionally, researchers probed for reasoning when participants gave limited responses. Interview questions were piloted with a QA manager within our own medical school. Very minor changes were required to enhance clarity, appropriateness and sense-making. All interviews were conducted one-to-one apart from two, where two people from the same organisation were present to provide comprehensive responses. Telephone/Skype interviews were conducted by four members of the team (AG, PC, LM, MP), audio-recorded and transcribed verbatim. Participants were geographically dispersed across the world therefore face-to-face interviews were not feasible.

Data analysis

Data were analysed following the stages of framework analysis.[17] This approach was followed due to its retroductive inductive-deductive nature to test initial theory while identifying emerging Context, Mechanism and Outcomes (CMOs) configurations. For *familiarisation*, two researchers each read one transcript and then made notes to identify CMO configurations (Total: 5 researchers × 5 transcripts). All researchers met to discuss similarities and differences across the transcripts including recurrent CMO configurations. We then developed a coding framework, consistently applied to each transcript. Five researchers coded data using NVivo,[18] with frequent progress meetings.

Reflexivity

Prior to analysis, individual members completed a written reflexive exercise which highlighted prior dispositions towards the research, and were then discussed collectively. The team consisted of practitioners, academics and researchers from medical and social science background disciplines, with QA knowledge ranging novice-expert.[19] Team members were vastly experienced in qualitative methods[20 21] and had previously applied a realist lens to understand complex interventions in healthcare education.[22]

Patient and public involvement

Given the focus on regulator-medical provider interactions in this study, patients and the public were not involved in the design, data collection or data analysis.

RESULTS
Participant details

Following ethical approval, we conducted interviews with 36 individuals representing 34 different organisations between July and September 2018 (table 1) to produce a considerable amount of data: 35 hours, 27 min. Interviews ranged between 48 and 88 min, mean=63. The sample represented regional, national (England, Scotland, Northern Ireland, Wales) and international stakeholders within and outside medicine (figure 3). There were 12 QAPs and 22 non-QAPs, with 27 (79%) of these from the UK and 7 (21%) international, representing Asia, North America and Europe. Participants often held senior roles such as chief executives and quality leads.

Main findings

We present findings which verify, refute and challenge our initial programme theory (themes 1–4), leading to the development of our modified programme theory. Contexts, mechanisms and outcomes are labelled as [C], [M] and [O], respectively. CMO configurations, resources and responses are identified and illustrated across themes.

Theme 1: quality standards

Standards defined the level at which a provider needs to function to reach certain outcomes, for example, meeting minimum standards. Key mechanisms triggered by co-construction of setting standards included compliance,

Table 1 Demographic information of the participant organisations

Interview number	ID	QA partner (Y/N)	Profession Sector	Location (coverage)
1	HealthMedINT1	N	Health, medicine	International
2	QAUG2	Y	Undergrad	UK
3	QARC3	Y	Royal College	UK
4	OtherprofessionUK4	N	Other Profession	UK
5	QAUG5	Y	Undergrad	UK
6	HealthMedUK6	N	Health, medicine	UK
7	HealthMedUK7	N	Health, medicine	UK
8	EducationUK8	N	Education	UK
9	HealthNon-MedUK9	N	Health, non-medicine	UK
10	QAPG10	Y	Postgrad	UK
11	HealthNon-MedINT11	N	Health, non-medicine	International
12	QAPG12	Y	Postgrad	UK
13	HealthMedINT13	N	Health, medicine	International
14	QAPG14	N	Health, medicine	UK
15	QAUG15	Y	Undergrad	UK
16	QAPG16	Y	Postgrad	UK
17	EducationUK17	N	Education	UK
18	HealthNon-MedUK18	N	Health, non-medicine	UK
19	QAUG19	Y	Undergrad	UK
20	HealthMedINT20	N	Health, medicine	International
21	HealthMedUK21	N	Health, medicine	UK
22	OtherprofessionUK22	N	Other profession	UK
23	HealthMedUK23	N	Health, medicine	UK
24	QARC24	Y	Royal College	UK
25	EducationUK25	N	Education	UK
26	HealthMedUK26	N	Health, medicine	UK
27	HealthNon-MedUK27	N	Health, non-medicine	UK
28	QAPG28	Y	Postgrad	UK
29	QAUG29	Y	Undergrad	UK
30	HealthMedINT30	N	Health, medicine	International
31	EducationUK31	N	Education	UK
32	EducationINT32	N	Education	International
33	QAPG33	Y	Postgrad	UK
34	HealthMedINT34	N	Health, medicine	International

Health, health organisation; INT, international based; Med, medical organisation; PG, postgraduate; QA, quality assurance partner; RC, Royal College; UG, undergraduate; UK, UK based.

clarification, flexibility and adherence. Undergraduate and postgraduate QAPs responded to the regulators standards by inclusion in their own policies. The standards also provided QAPs with leverage to push forward changes at institutions:

I think, a very positive element [O], is that it [standards] has allowed UK medical schools [C] within the framework to differ in how they implement that framework…So I think as well as having the rigour of what must be done, it allows for a degree of flexibility [M]. (HealthMedINT1)

We reference them [the standards] in our…internal policies. (QAUG19)

However, the standards had unintended consequences as their presence sometimes created confusion, particularly in the postgraduate context due to lack of clarity. In some instances, organisations had their own standards to assess educational quality, resulting in confusion:
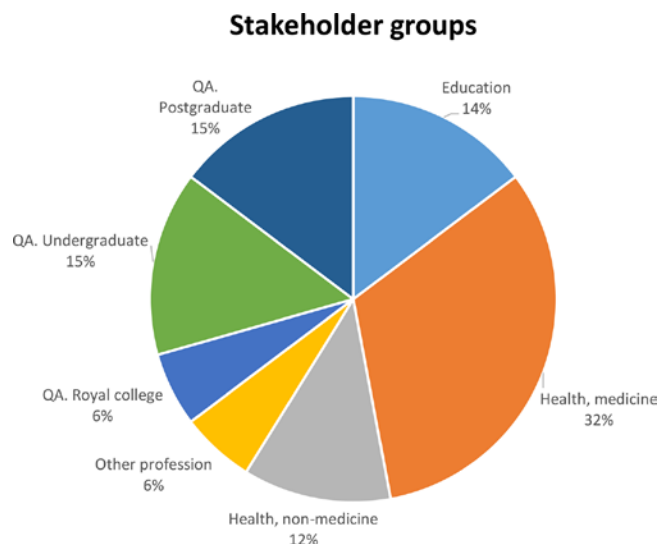
Stakeholder groups

Figure 3 Stakeholder groups. QA, quality assurance.

I suspect [postgraduate organisation] ignores them [standards] because they've come up with their own quality framework. (QAPG33)

Our biggest concerns [O] really are not so much the standards as the sort of processes that by which the GMC [C] will check that our curricula are… comply with those standards [M]. And I think on that, you know, looking back now I can see that there was a certain ambiguity [M] in how the GMC were going to approach this and I'm not sure that they ever resolved it as the standards were being developed [O]. (QARC3)

Prescriptive, rigid and inflexible standards prevented providers from being adaptable to need and innovation. For example, standards which focus on particular aspects (eg, student diversity) detracted attention from other areas of need (eg, widening participation). Conversely, less binding standards (eg, not detailing specific teaching methods) triggered mechanisms of ambiguity, openness and flexibility creating too much variation in education across contexts and producing new risks to quality. Misalignment between different quality standards caused frustration. For example, a LEP was deemed to be clinically outstanding but was also found to be inadequate for educational quality by a different regulator. Local pressures were seen to inhibit postgraduate partners' abilities to follow standards, suggesting that in the 'real world', applicability of standards was sometimes questionable:

A lot of LEPs take our students [C], but they [LEPs] can quite readily tell us [medical school] to take them away as well [O], if we're very strict with them about meeting certain standards and certain criteria [M]. (QAUG5)

Theme 2: sanctions and approvals
We identified that organisational culture affects approaches to sanctions, and so an 'acceptable sanction'

was contingent on risk. However in different contexts, should supportive measures be inadequate then the most severe sanction of withdrawn approval should remain. The 'ultimate sanction of power' (QAPG28) was regarded to fulfil its intended consequence, that is, to protect patient safety, but also had unintended consequences to reinforce the medical regulator's authority and subsequently motivate providers to address problems. There was a firm belief that a severe sanction should rarely need to be enforced if other QA components (eg, self-assessment) are effective.

It's [sanctions] a bit of a lightning rod situation, but I think it [closing medical school] should remain as the ultimate sanction [O]…If trust management realised for example that they wouldn't lose their trainees as a result of not providing a safe and effective training environment [M]… I think [it would] slip further down their list of priorities [O]. (QARC24)

The effectiveness of regulatory approvals in the undergraduate and postgraduate context varied. In undergraduate contexts, it was described as time-consuming examining both curriculum and staff capabilities. However, in the postgraduate context the process and need for QA was not as clear. Non-QAPs felt that it was important that mechanisms were in place to periodically review approvals. For instance:

[The thoroughness] enabled them [regulator] to make a decision on our suitability to proceed. (QAUG29)

what if the trainee goes for 1 week, but it's only 1 week out of a 1 year placement, do they need to get that site approved? (QAPG12)

We don't link approvals and quality very strongly [C]…we go to the GMC and we say, can we put some doctors here please? And the GMC go, yes. But there's an implication in doing that that because we're asking, we're going to quality manage that particular set of placements [M]. And we do, but not explicitly and not formally [O]. (QAPG16)

Theme 3: collecting information: visits, monitoring and self-assessment
Institutional visits were positioned as a key component as they triggered internal reviews and reflection, subsequently motivating organisations to improve quality. Working collaboratively engendered trust with open and honest dialogue which was considered crucial in effecting change. Meaningful dialogue and collaboration were important and that was achieved through having high quality, 'respected' (EducationINT32) trained visitor teams:

I was prepared to be completely open and honest with the GMC…If [visits] are going to be effective, relationship building is actually more important than what you're doing collecting evidence. (HealthMedUK21)

The QAF includes a range of monitoring data collection processes such as: data from the National Training Survey (NTS); monitoring including enhanced monitoring; self-assessment; and visits. The NTS surveys all doctors in training which facilitated increased transparency, accountability and risk-identification. The resources provided by the survey could lead to invaluable outcomes to examine training differences, make evidence-informed decisions and pinpoint training issues. However, multiple sources of data were sometimes regarded as conflictual, obscuring the overall picture of education quality.

> At the moment, they're [GMC] looking [at] trainee burnout [C]. So they're generating all this data at the moment and I don't think they're clear about what they're going to do with it [M], and my concern is they will just dump it on us for us to fix, and I don't think we can [O]. (QAPG10)

The component of requiring self-assessment triggered many different mechanisms in different contexts. For regulators, it generated reflective internally-led processes. The reasons for this were around connectivity between regulators and providers:

> [self-assessment] a really fundamental part of what we do, and we place a massive… emphasis on that. (EducationUK17)
>
> this is the way [self-assessment] an institution connects itself with given standards (EducationINT32)
>
> to work constructively with the provider…being the start of a peer review process (HealthNon-MedINT11)

Whereas, for those who were being quality assured, the formality of written self-assessment inhibited open disclosure as it was laborious and seen more as an audit. Validity and reliability was also raised as perceptions of lack of feedback from the regulator also undermined self-assessment.

> I think you're more likely to hear genuine issues, genuine things that need to be fixed, if you speak to people informally and off the record (QAPG6)
>
> It [self-assessment] forms part of it[assuring quality] and it's a very strong part of it, but I wouldn't necessarily use it [self-assessment] in isolation (HealthMedUK23)
>
> We are encouraging of institutions identifying challenges [C]. So if an institution is very open and honest [M], even into what might be quite a delicate area, saying this has been a challenge for us [C] and we're working away on it and we're doing the following things. Provided that their plan of action is a good one and that it's being conducted in a timely manner that would be reported on in a positive light [O]. (HealthNonMedUK9)

Theme 4: reporting—accountability, dissemination, good practice

A patient safety outcome response identified from external reporting was to build public confidence. Publicly available outputs fostered accountability to illustrate how providers are low-risk thus requiring less scrutiny. Insufficient reporting and feedback (in terms of timeliness, quantity and quality) fostered outcomes of devaluing time and effort, and subsequent disengagement. Risk context was also important to determine the effectiveness of intervention components. Informal partnerships were highlighted as critical to assuring quality.

> I think the transparency in publications are important because it involves or it makes things clear and open to all stakeholders. (QAPG28)
>
> I think what having it public does, is it creates some pressure and accountability [M] on both the accreditor and the accrediting body [C] to focus on the outcomes and to show progress against conditions [O]. (HealthMedINT13)
>
> Working in partnership with regulators was instrumental, it has a significant effect on driving change in trusts (HealthedUK21)

When feedback mechanisms triggered included collaboration and openness, this fostered informal working partnerships leading to a raft of positive outcomes including awareness, sharing knowledge and quality improvement. Rapport over time helped provider's develop trust to report concerns. Here, the positioning of regulator-provider context, moved from accountability quality-checker to collaborative problem-solver. Spreading good practice was contextually variable.

> The institution needs to take that genuine look at it, and spend the time genuinely evaluating and genuinely creating action plans…an institution that is good at critical self-reflection will tend to address problems before, or potential problems, before they become actual problems [O]. (EducationUK17)
>
> What works for one school may not work for other [O] …So, you don't want people to blindly be saying oh, let's do that, because that's going to be good practice here [M]… Because education in programs do vary [C] (HealthNon-MedUK18)

Subsequently, motivation reaction was low to implement changes based on other organisations' examples of good practice. Reports were deemed most helpful when they included action plans accessible to lay audiences.

DISCUSSION
Summary of findings

We undertook a realist evaluation to explore UK medical education QA. We found that intervention components support or undermine QA for different organisations, and at different times in undergraduate (for medical students) and postgraduate (for trainees) contexts. In

the results we revealed that although interventions were often implemented uniformly in undergraduate and post-graduate contexts, the impact of these varied with some leading to positive and negative outcomes. We tested our initial programme theory (themes 1–4) to develop a modified programme theory.

Across the three-tier model we identified that the undergraduate and postgraduate context were influential. For example, in the undergraduate context, the leverage brokered by the regulator was often associated with directive features enabling local changes. Whereas in the postgraduate context, this power was often lost and diffused across education layers. Predominantly in the postgraduate trainee context, interventions led to unintended consequences (eg, organisation disengagement) if an intervention promoted adherence at the cost of autonomy, subsequently triggering a lack of motivation (theme 1). An underlying mechanism identified to ensure an inclusive approach to QA was partnership working (themes 3–4). In the undergraduate context, provider-regulator engagement was sometimes not present, typically when there was a lack of informal relationships. In theme 3, visits were identified as a component that could better foster partnerships; so long as they were conducted with integrity, meaning and purpose.

Collectively across themes 1–4, the QAF was commended for being comprehensive and enabling a broad understanding of an organisation's performance. Internally-led processes with organisations identifying and addressing their own challenges and deficiencies, when done well, promoted a sense of autonomy and accountability (theme

3). The main unintended QAF consequences fell broadly into two outcomes, those related to the overlap across the three tiers (themes 1–2) and those related to the regulatory burden associated with data-driven approaches with a lack of transparency on why and how data were used (themes 3–4). A blurring of roles and boundaries of multiple organisations between patient safety, medical education and training was identified.

The influence and power of the regulator was continually picked up across the components (themes 1–4) which triggered mechanisms including transparent reporting to promote quality improvement, effective communication, trust and partnership working facilitating interactions between regulators, partners and providers. Proportionate reactions in the face of disclosing and identifying patient safety risks at an early stage were more likely to occur within a positive trusting regulator-provider context underpinned by openness. Likewise, an organisation that self-assessed critically was reported to give regulators confidence in the institution.

Modified programme theory

The findings informed a modified programme theory to explain the underlying processes for the intended and unintended consequences of how the GMC quality assures education in various contexts (figure 4). The modified programme theory conceptualises CMO configurations presented so that components are understood in the ways in which they may lead to positive or negative outcomes. The programme theory broadly conceptualises the differing QAP contexts as having associative
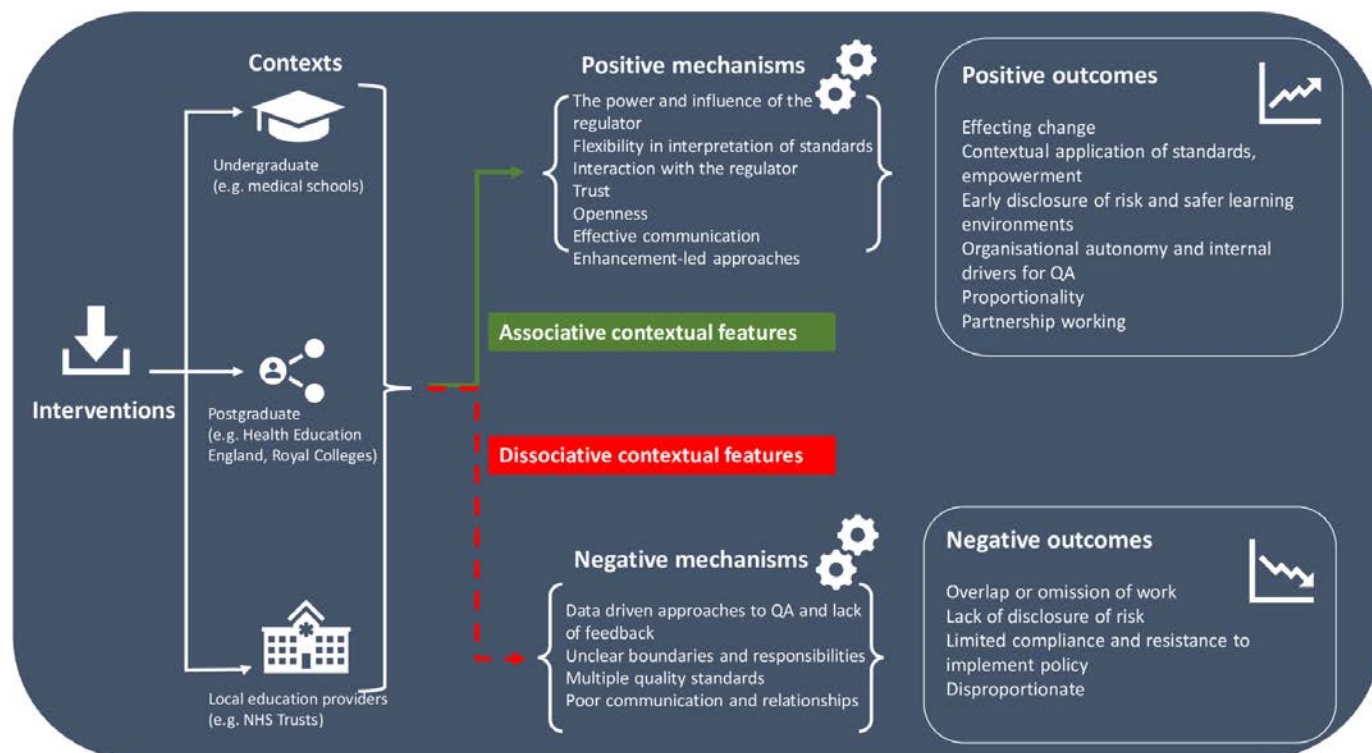

Figure 4 Modified programme theory to explain the underlying process for the intended and unintended consequences of QA components. NHS, National Health Service; QA, quality assurance.

or dissociate contextual features. Associative contextual features are exhibited where the QAP demonstrates adherence to the regulator's QAF, either in full or to particular components, triggering positive mechanisms and outcomes. Dissociative contextual features are where there is organisational resistance to imposed external QA triggering negative mechanisms and outcomes. Each of the QAF components can therefore be enacted differently within these different contexts. Across the three-tier model (QA, quality management and quality control), the theory demonstrates that QAF components are enacted differently depending on whether the context has associative or dissociative features, with implications for the mechanisms triggered leading to positive (eg, effecting change, contextual application of standards, partnerships) and/or negative outcomes (limited compliance, resistance, overlap).

Relevance of findings and implications

The findings reinforce the QA literature highlighting *trust* in fostering effective working relationships to enhance feedback.[1 5] We extend this further, and identified that early communication of emerging risks supports quality assurance and enhancement approaches through informal networks. Visits aid communication and build relationships, yet if lost may distance the regulator and undermine opportunities for partnership working. Informal communication provided a safe environment for providers to discuss concerns with the regulator, opposed to formal monitoring acting as a barrier.

Expanding the literature, we demonstrate that context must be considered in order for quality assurance to protect patients.[23] Risk is context dependent and was perceived to be tangibly different across undergraduate and postgraduate contexts. Undergraduate medical settings were perceived as low risk and imply opportunities for greater tailoring and focus. The overlaps between quality assurance, quality management and quality control were apparent especially within the postgraduate setting with duplication and confusion of responsibilities. These findings align with a recent systematic review identifying features of failing healthcare organisations including conflicting missions, fragmented accountability and lack of collaboration.[1]

Collectively, this supports the need to clarify structural quality processes and how organisations are intended to function collaboratively. In the analysis, risk-based visiting positioned the regulator as quality assurer rather than quality enhancer. Equally, effective assurance is often associated with suppressing innovation.[24] Moreover, the role of self-assessment[24] posed a number of challenges in relation to purpose and autonomy. While institutional self-assessments can positively influence reactions to drive quality improvement, there are issues with validity, reliability and internal quality review.[19 24–27]

The power of the regulator impacted on the effectiveness of intervention components in multiple ways. The regulatory-burden associated with monitoring activities was considerable and disengagement ensued. Lack of feedback from the regulator was an important aetiological mechanism precipitating the situation. Similarly, negative consequences of approvals including cost, low staff morale, threats to organisation reputation and the suppression of innovation through adhering to standards has been identified.[28 29] Without regulators addressing varying risk contexts, the proportionality of QA is imbalanced, leading to negative outcomes with regulators unable to effectively assure quality. Therefore, collectively considering a hybrid model of cyclical plus risk-based visiting may help to build provider relationships and drive improvement while ensuring minimum standards. Collective assurance and relationships should be encouraged so that regulators and providers can tackle issues conjointly. Flexibility in utilising other datasets within any collaborative work is a necessity and a clear stance on organisational remit and particularly boundaries, is anticipated to be a key mechanism in effective joint QA.

Strengths, limitations and future directions

To our knowledge, this is the first robust study on education QA within the healthcare context, synthesising data from stakeholders. The study fills a gap as QA remains expensive, yet its functionality is largely unexplored. The study was conducted by an experienced multidisciplinary research team applying an innovative realist approach, underpinned by a sound team-based analysis. A somewhat surprising omission from our findings is a lack of attention to the mechanism of leadership.[1 23] The sample focused on processes rather than delivery perhaps contributing to such omission. Moving forward, there is a need to conduct an economic review and consult with stakeholders into what data could be shared (eg, NTS, Care Quality Commission data) to understand links to intervention components. The findings have influenced the GMC's approach to QA which impacts on all medical students and doctors in training.[30]

Conclusions

This study used a realist methodology to reveal the intended and unintended consequences of components used by the GMC to quality assure medical education, and elucidated the mechanisms by which both were brought about. While uniform approaches are often in place, interventions need to be contextually tailored. Continuous partnership working can enhance open disclosure to drive up education quality. This research has provided a modified programme theory to explicate how education providers and regulators can work more effectively together to uphold quality, and ultimately protect public safety.

ORCID iDs
Paul Crampton http://orcid.org/0000-0001-8744-930X
Michael Page http://orcid.org/0000-0003-1932-9925

REFERENCES

1 Vaughn VM, Saint S, Krein SL, *et al*. Characteristics of healthcare organisations struggling to improve quality: results from a systematic review of qualitative studies. *BMJ Qual Saf* 2019;28:74–84.

2 General Medical Council. Promoting excellence, 2019. Available: https://www.gmc-uk.org/education/standards-guidance-and-curricula/standards-and-outcomes/promoting-excellence [Accessed 16 Jul 2019].

3 Vroeijenstijn AI. Quality assurance in medical education. *Acad Med* 1995;70:S59–67.

4 General Medical Council. How we assure quality, 2019. Available: https://www.gmc-uk.org/education/how-we-quality-assure [Accessed 10 Apr 2019].

5 Jones K, Tymms P. Ofsted's role in promoting school improvement: the mechanisms of the school inspection system in England. *Oxf Rev Educ* 2014;40:315–30.

6 Westerheijden DF, Stensaker B, Rosa MJ, eds. *Quality assurance in higher education: Trends in regulation, translation and transformation*. Vol. 20. Springer Science & Business Media, 2007.

7 Haji F, Morin M-P, Parker K. Rethinking programme evaluation in health professions education: beyond 'did it work?'. *Med Educ* 2013;47:342–51.

8 Øvretveit J, Gustafson D. Evaluation of quality improvement programmes. *BMJ Qual Saf* 2002;11:270–5.

9 Pawson R, Greenhalgh T, Harvey G, *et al*. Realist review--a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10 Suppl 1:21–34.

10 Pawson R. *The science of evaluation: a realist manifesto*. London: SAGE Publications Ltd, 2013.

11 Jones AC, Shipman SA, Ogrinc G. Key characteristics of successful quality improvement curricula in physician education: a realist review. *BMJ Qual Saf* 2015;24:77–88.

12 Wong G, Westhorp G, Manzano A, *et al*. RAMESES II reporting standards for realist evaluations. *BMC Med* 2016;14:96.

13 Griffin A, Crampton P, Mehdizadeh L, *et al*. Understanding stakeholder perspectives on the GMC's quality assurance of medical education and training. Final Report, 2018from. Available: https://www.gmc-uk.org/-/media/documents/understanding-stakeholder-perspectives-on-the-gmcs-quality-assurance-of-medical-education-and-traini.pdf?la=en&hash=14529235D9E6C99AF9B5EDA82F6848CB34D33396 [Accessed 25 Jul 2019].

14 Jagosh J, Bush PL, Salsberg J, *et al*. A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. *BMC Public Health* 2015;15:725.

15 Manzano A. The craft of interviewing in realist evaluation. *Evaluation* 2016;22:342–60.

16 Keren G, De Bruin WB. On the assessment of decision quality: Considerations regarding utility, conflict and accountability. In: *Thinking: psychological perspectives on Reasoning, judgment and decision making*, 2003: 347.

17 Ritchie J, Spencer L. Qualitative data analysis for applied policy research. *The qualitative researcher's companion* 2002;573:305–29.

18 QSR International Pty Ltd. *NVivo qualitative data analysis software*, 2015.

19 Griffin A, McKeown A, Viney R, *et al*. Revalidation and quality assurance: the application of the MUSIQ framework in independent verification visits to healthcare organisations. *BMJ Open* 2017;7:e014121.

20 Griffin A, Knight L, McKeown A, *et al*. A postgraduate curriculum for integrated care: a qualitative exploration of trainee paediatricians and general practitioners' experiences. *BMC Med Educ* 2019;19:8.

21 Rees CE, Crampton P, Kent F, *et al*. Understanding students' and clinicians' experiences of informal interprofessional workplace learning: an Australian qualitative study. *BMJ Open* 2018;8:e021238.

22 Ajjawi R, Crampton PES, Rees CE. What really matters for successful research environments? A realist synthesis. *Med Educ* 2018;52:936–50.

23 Kaplan HC, Provost LP, Froehle CM, *et al*. The model for understanding success in quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf* 2012;21:13–20.

24 Lang DW. Self-Regulation with rules: lessons learned from a new quality assurance process for Ontario. *Quality Assurance in Education* 2015;23:216–32.

25 Tackett S, Grant J, Mmari K. Designing an evaluation framework for WFME basic standards for medical education. *Medical Teacher* 2016;38:291–6.

26 Jones KL, Tymms P, Kemethofer D, *et al*. The unintended consequences of school inspection: the prevalence of inspection side-effects in Austria, the Czech Republic, England, Ireland, the Netherlands, Sweden, and Switzerland. *Oxf Rev Educ* 2017;43:805–22.

27 Forrest C. Inspection and improvement in three further education colleges. *Research in Post-Compulsory Education* 2015;20:296–314.

28 de Paor C. The contribution of professional accreditation to quality assurance in higher education. *Quality in Higher Education* 2016;22:228–41.

29 Blouin D, Tekian A, Kamin C, *et al*. The impact of accreditation on medical schools' processes. *Med Educ* 2018;52:182–91.

30 General Medical Council. Council papers, 2019. Available: https://www.gmc-uk.org/-/media/documents/council-agenda-and-papers-for-meeting-on-27-february-2019_pdf-78924314.pdf [Accessed 17 Jul 2019].