**Listen to THIS - Episode two: Learning from failure in healthcare**

Graham Martin: Welcome to Listen to THIS, where we have conversations examining how we can improve the quality and safety of healthcare.

In this series we'll dive into both current and longstanding healthcare challenges and will shine a light on some of the work that's being done to address them.

I'm Graham Martin, Director of Research at the Healthcare Improvement Studies Institute.

Learning from mistakes is a crucial part of improvement and as humans we tend to focus on the negatives. But if we concentrate on just the mistakes, are we actually hindering progress?

In this episode we'll focus on an important subject: learning from failure. We'll explore the positive bias in quality improvement, the differences in academic research and service investigations, and the valuable insights we can get both from when things go right and when things go wrong. I'm joined today by…

Jane O’Hara: I’m Jane O’Hara, and I’m a Professor and Director of Research based at the Healthcare Improvement Institute which is based within the University of Cambridge.

Helen Crump: Hello, I'm Helen Crump, I'm a THIS Institute Research Fellow at the University of Birmingham's Health Services Management Centre, researching how unsuccessful quality improvement projects and processes taking place in healthcare are reported and shared.

James McGowan: Hello everyone, I'm James McGowan, I'm a public health physician, I'm currently a consultant in Health Protection at the UK Health Security Agency and I'm a former fellow at THIS Institute where I led work that focused on building evidence to support challenged healthcare organisations and clinical services.

Graham: And thank you all three of you for joining us. I suppose the main thing that triggered the idea for this conversation was something about almost the contrast between Helen's PhD – which as she was saying in the introduction – focuses on improvement and the way in which improvement tends to accentuate the positive and leave out all the negative.  And James’ PhD, which focuses much more, as he said, on learning from failure.

But when were meeting to discuss this beforehand, I think something that really struck us quite strongly was that actually there is something of a difference between the way that failure is treated in these different approaches to gaining knowledge about healthcare.

Research on the one hand tends to almost always…or very often…focuses on what's gone wrong because that's so much easier to write about as an academic. If everything's fine, then what's the story?

Whereas as Helen has been finding, and I'm sure we'll hear more about this later on the episode, it seems to be true in health improvement work.

Jane, I wonder if we could bring you in on this first and get your reflections on that, because I think you pointed out that apparent tension when we were talking about this beforehand.

Jane: I think it's interesting and I think I've since reflected on the fact that I'm not sure that the split is entirely between QI and research.

In that sense, I think the biggest split is in between how within the healthcare community we often focus much more on failure, much less on things either going well or indeed just how things go, and how we might improve system performance or outcomes for people.

But coming back to your point, I think that, as I say, I've reflected and there was quite a lot of positivity bias in research as well. I mean, I know sociologists want to describe the problem oftentimes and have been criticised for that, but I think certainly in interventional quality and safety research the same problems occur. There's a publication bias in reporting of positive findings.

It's a lot harder to publish negative findings or negative trials. And a lot of work has gone into that in terms…particularly in terms of interventional trials for drugs and different things.

But I'm not sure that we've necessarily caught up entirely in the sort of healthcare improvement world.

So I think broadly the split is much more between the healthcare community and how it looks to improve…and identification of the sources of things that need improving within healthcare often comes from failure and the sort of positivity bias which I think you do see very much.

And Helen's going to talk about her work on that, in presenting what works rather than actually learning from all kinds of activity where we're trying to improve the healthcare system.

Graham: Thank you, Jane. Helen, I'd love to hear your thoughts on that. James mentions publication bias there and I think I might be right in thinking that was one of the bits of background knowledge, that tendency to publish positive findings and under publish either null or negative findings from things like pharmaceutical trials, and the extent to which that also manifests in health services research. I think that was one of the things that originally prompted your study, wasn't it?

Helen: That's right, yes. So one of my supervisors actually had done some research looking at this area and I think identified signs of a positivity bias in health services research as well. And I think that definitely does exist. I think we might come to talk about this in a little bit more detail shortly.

But one of the interesting things about quality improvement is that often quality improvement projects arise in a slightly different way from something that might be viewed as more conventional academic led research project.

So we don't necessarily have things like protocols that help us to assess performance against an original set of objectives.

Helen: So I think that's one area where maybe I don't want to say it's easier not to report failure in quality improvement, but it's perhaps one area where we just have less of an opportunity to be forced to talk about things that haven't worked well.

The other side of this is we see much more of a preponderance of grey literature publications in quality improvement. Often they're great, they're bringing information to people quickly, so much more rapid and accessible way of sharing information.

But they do tend to skew towards best practice, even though people often acknowledge privately that they learn much more from things that haven't quite worked out than from straightforward successes.

So I think there are a few reasons for it, something that we need to consider when thinking about how we talk about quality improvement outcomes and projects.

Graham: That's really interesting and I think you pointed towards the multifactorial influences on this positivity bias, such as it is in quality improvement, if you had to say what is really at the core of that?

Helen: I think there are two aspects to this. I think one is the practical side of it, which I've already alluded to a little bit, and the other is the human side of it. So, on the practical side of things, on the face of it, we should be able to learn from failure in quality improvement in similar ways to what we do in terms of patient safety, I guess.

But in practice we don't have the same kind of infrastructure to help with that in quality improvement. I mean, quite understandably, we don't have anything that works in quite the same way as ‘Never Event’ reporting. And as I said, quality improvement works often rather pragmatic. So a project might start when someone finds a practical problem that they want to address.

Hopefully there'll be some kind of measurement of the baseline position and some clearly articulated aims so that we can see whether what's being tried is working.

But because of the nature of quality improvement, it's possible, particularly if PDSA (plan-do-study-act) cycles are being used, that a project will go through multiple iterations before it's completed. And I think that can make it quite tricky when you're trying to define and describe failure in relation to these kinds of projects.

So say the failure part happened during an early PDSA cycle and the issue was then corrected, leading to a successful outcome. Do the project team still consider that a failure or is it actually a success? So I think there's a fundamental issue around actually identifying a failure.

And then so the challenge then becomes to identify these failures and extract the learning that's generated through the experience of something not working out initially, but possibly working out better in the end. And then, of course, there are projects where the intervention just doesn't work, where it's a bit more of an obvious, straightforward failure.

And I think that's where the human side of this comes in. People just don't like to talk about things that haven't worked out. And outside of patient safety, there's less of a culture of talking about failure.

And so it's easy to just put it to one side and move on. I think people get worried, understandably, about whether talking about something that's failed might have negative repercussions, like perhaps for their career or for their organisation's reputation.

And so if we want people to be able to use and share learning from failure to a fuller extent, I think we just need to make it feel safe for them to be able to talk about this stuff. So I think that's the big challenge that we have to crack if you want to make this kind of information from failure a bit more accessible.

Graham: And I think there's something really interesting there about the way that the good reasons for doing something might align with the bad reasons for doing the same thing.

And actually picking those apart and deciding what's best is quite difficult to do, because clearly, if you're going to report something in a way that people read, you're not going to show every single item of dirty laundry and list them one by one. You really want to get to what really matters.

But at the same time, there's an incentive to just focus on the positive stuff at the end and leave out all the learning on the way. So I think that's quite tricky to do, really.
Jane, I think you might want to come in there.

Jane: I think it's really important to differentiate between learning from failure in a quality or safety way and learning from the failure of an intervention or an improvement effort designed to improve that.

And I think one of the things that QI has come out of potentially, is the fact that there's a will to improve the quality and safety of healthcare. With clinicians, people want to do their very best. They get frustrated with systems, they want to change them. They feel that they're very close to those systems, so they know what the problems might be and the potential ways of changing them.

But often the prompts for those problems are failures. So the inputs into the system are things that have been assigned by the system as being failures or things that have gone wrong.

Jane: And so, in a sense, you wonder whether that…I don't like to use this term, but in the sense of toxic positivity that we're talking about with respect to QI…might be a real reflection of the fact that it is the time where people working in health services can actually take control and say, “okay, we're just going to try and make this better and we're going to not just focus on the negative”, which, as you say, the world of investigations, enquiries, patient safety initiatives can sometimes feel quite negative.

And QI feels like it occupies a very different space. So I imagine that there's, as you say, there's multiple reasons, Helen, why things don't happen.

The other thing, and I know James is going to come in soon, but the other thing is these aren't done as research projects. We are, as researchers and academics, we are…that's our bread and butter.

We say what we're going to do, we do it and then we report on it. If your intention is to try and improve the service, you think, well, I didn't improve service, so I don't need to tell about it…and I'm paraphrasing there, I'm probably being a bit reductive, but we would see that as really important learning.

But is there a sense within the community that is such important learning? If you haven't demonstrably improved anything, are they going to reflect on the processes of the improvement in quite the same way?

Helen: I think Jane makes a fundamentally important point there, which is this point about these projects, coming up in quite an organic way, and they're not necessarily being the route to actually report findings without really thinking about it quite consciously, I think.

And often you talk to people who are working in these areas, they aren't researchers as such, they're generally clinicians or managers and they've identified a problem and done some stuff to try and fix it. That stuff may or may not work, and something may not have worked initially, but then worked in the end. And often they don't have a framework to process to share that information, either within their organisation or externally.

So when you talk to people, they say, oh, yeah, the learning from stuff that hasn't worked is at least as valuable as the learning from stuff that has worked. But often that's where it stops. And there isn't the focus on sharing this, boiling it down into what's the really important key learning points that we're going to take away from this and how are we going to do things differently in future?

And when it does happen, it's often quite tacit knowledge that's being shared. So people talking to each other in the workplace, not recording stuff or saving it. I mean, that may not be a problem. It's just a consideration when we're thinking about how we accept this knowledge and how we make sure that we can use it to its fullest potential.

Graham: James, would you like to come in here?

James: Just to extend the point that Helen, Jane, I think, have both made. So underlying this is, you have this gap between the approach to evidence and evaluation, between the sort of healthcare community practitioners and how researchers deal with those two. And I just wondered whether either of you had any reflections on how to close those gaps between the practitioner and the researcher community.

We'll come on to my project in a minute, but part of my project has been reviewing the maternity improvement literature. One of the things that strikes you about it immediately is just the dominance of grey literature. So an awful lot of what we know about maternity improvement is documented, reported in grey literature sources, not necessarily in formal academic journals. That's what we have as a corpus of knowledge to go on.

So I think that illustrates the point in terms of the gap between the two communities. So I wonder whether either of you had any practical reflections on how to address that.

Jane: I think it's a basic - that's an incredibly important finding from your work that actually most of what's happening to improve maternity services is happening outside of the research literature. And that has implications for the evidence base, it has implications for the methods that are used.

And I think the visibility of work and the visibility of the focus on maternity at the moment is both a blessing and a curse in the sense that it's highlighted that things need to be done that…actually there's a whole improvement agenda, academic agenda, that is there to be done over the coming months and years. But I think there's also then a push for improvement work to be done at pace.

I think the real frustration, of course, with research activity is just the delay and the time that it takes to do anything and to get evidence out into practice. And I think, if I may be controversial, there's sometimes a suspicion in clinical circles about academics doing work that directly speaks to healthcare improvement. I think that sometimes there's a sense that we're very removed from the realities of delivering care and I think to some extent in the past particularly that has been true.

So I think for me, the only way that we can move forward would be to try to bring communities together to say what questions need to be answered academically, what can be done more quickly within improvement work, but particularly in taking on board what Helen's work has been focusing on, how can we make the improvement work more robust across the totality of that work?

Particularly reflecting on methods, particularly thinking about the validity of those methods and the fidelity to those methods, thinking about all the lovely papers which basically say that 78% of whatever it is, PDSAs barely get past P or D or whatever it is. I think there's…particularly in maternity…I think there's a real opportunity, I think, to bring these two groups of people together to do it differently and better.

Graham: Helen, what do you think?

Helen: I think it's such an interesting question. I'm very aware that in terms of the projects I've been looking at, they're very much at the pragmatic, local end of the spectrum.

So I think the vast majority of the projects that I've been talking to people about, they probably wouldn't ever result in…I certainly don't think they would result in a peer reviewed journal article. I think they probably wouldn't result in a case study or grey literature or something like that.

Most of the time they might do if someone found it interesting and happened upon it. And they aren't really designed to fulfil that purpose. It's about trying to achieve an improvement.

So when you think then about how to extract information from something like that and share it's a very different set of considerations from something like doing a systematic review of literature or something.

I mean, I tried to do that at the start of my PhD and it did not work. So I can vouch for that not being the way forward on based on personal experience. What I'm interested in as part of my research is seeing if there is a way to help people to boil this stuff down and where a quality improvement project hasn't work, possibly to identify, was it one of these things that caused a problem.

So I've been looking at where I can find examples of things that haven't worked being talked about in the literature. Now, as I've already alluded to, that's quite hard because there isn't a great deal of stuff that's published.

You can sometimes find it in reports of interventions or projects where it's ultimately been successful and people are talking about a kind of interim failure. But I think you can start to see a bit of a map of issues.

So sometimes there's an issue with the quality improvement intervention itself, sometimes there's an issue with the way that was implemented. Sometimes it's to do with the organisational system context in which it was implemented, sometimes it's to do with the way that the performance of the intervention was measured.

And I'm wondering whether that's a starting point for starting to flesh out a way to help people to talk about this where it's not being reported. But I don't know how helpful people find that yet. So that's something I'm quite keen to test.

Graham: I think it will be fascinating to see what comes out of that work. Helen.

So if we go back to James's question of how you bridge the gap between research and practice communities, and we know that there's umpteen systematic reviews published every week, we know it's really hard for people in the healthcare system to keep up to date with research evidence, and we also know, actually, that a lot of academic literature is quite staid, quite difficult to read.

It's not exactly what you want to be a reading of an evening, unless you're actually trying to help yourself get to sleep. And what people like Helen Bevan, for example, would say is that what persuades people is stories.

And I wonder, actually, Helen, just Helen Crump…listening to what you've just been saying there. There's something about the failure…Just talking about a failure and leaving it at that probably isn't especially persuasive, it isn't especially compelling. It's not a great narrative structure, is it?

So actually, there's something to be said for this is what happens, this is what we plan to do. We did a P, we did an A, we got a D, we got lost, we didn't get onto the S and the A, we had all of these difficulties, but then this is how we resolve them.

Helen: Definitely. One thing that crops up with this is people are much more comfortable talking about a failure where it's got a happy ending and they can talk about the fact that they overcame it and they did X, Y and Z, that fixed the problem, and then now everything's working well.

I suppose the problem with that is, in reality, that doesn't always happen and we still need to know about when something just doesn't work or we're potentially doomed to keep making the same mistake over and over again.

So I think on one level, there's this really thorny challenge that we have, which is almost about de-risking talking about failure. So trying to find a way to enable people to describe stuff that hasn't worked without feeling that it is potentially going to harm their career or bring their organisation into disrepute or something like that.

I think that's a really tough challenge and we can't be naive about it. But to enable people to have those conversations where there isn't the happy ending in sight, I think we really do have to help people to process this information in such a way that they can speak about it from a position of safety. Whether that's something like anonymising it, or sharing.

We could have workshops where we focus on a particular service area and get people to submit anonymised examples of stuff that hasn't worked out and then talk about those in that way.

It'd be interesting to explore what people might feel able to share and what contexts in which they might feel able to share it. But I don't think we can just dive in and say, let's all talk about our failures and it'll all be fine and everyone will be okay with that. I just don't think that's the world that we're in, unfortunately.

Graham: And as a quick aside, I think something that brings to mind for me is some debates that are ongoing in academia at the moment about retractions. So a paper is submitted to a journal, it passes peer review, it gets published, and then sometime later, someone, either the investigators or a third party, unearths an issue in the analysis or the data collection or something that basically undermines the validity of that research paper.

The right thing to do is to go for a retraction, but at the moment there's stigmatisation for fairly obvious reasons associated with retraction, because it suggests that the researcher wasn't doing their job properly. It means they've got one fewer publication, it could harm their prospects.

And so there's something…you use the word de-risking…and it strikes me that those kind of conflicts of interests between the individual and the collective that they're not uncommon. So there's probably things to be learned from various fields there in terms of what de risking could look like in practise in a way that makes it…aligns incentives a bit better, I suppose. Another thing you just mentioned there, Helen, was around the usefulness of the story of an individual, the dramatic line of: “we tried this, it went wrong, but then we heroically resolved it” and how transferable that might be to other places.

That brings to mind some of James’ research, and particularly the research that he's doing that focuses on the particular issues that face some kinds of services over the others, those that are often called challenged or troubled, the services that are just really struggling to get themselves out of the doldrums.

So, James, I'd love to hear a little bit more about your study, please.

James: Thanks, Graham. And really interesting listening to Helen, your reflections on your project. I think there's a lot of overlap and similarities. So the project I've been leading has been focused on, as Graham said, challenged NHS organisations in clinical services. So thinking about whole trusts or whole departments or specialties, for example.

So I have a few observations to share first, I think immediately you come up against the challenge of the observability problem of challenge status or the detection problem of failure. For the purpose of my PhD, I've focused on trusts and services rated as either inadequate or requires improvement by the Care Quality Commission, who are the regulator of trusts.

But obviously there are limitations for that approach and there's the sensitivity of these sort of formal systems of regulation and detecting struggling trusts have been challenged, including their relationship with performance indicators.

So obviously I needed an approach for my project to identify trusts and services that might be in trouble. But there are limitations to different approaches to doing so.

The second thing I would say is building on some of what Helen was saying, I think often the focus of people in the policy making community, certainly the NHS, I imagine sustainment healthcare systems, other healthcare systems, is about interventions to stimulate performance improvement.

So the quality and safety side, for example. But what we see in maternity services, which has been the focus of my project, is we see deficits in how these interventions are designed, implemented and delivered.

Part of the work I was leading was looking at large-scale improvement programmes and maternity services. So I looked at initiatives have been implemented in services going back as far as 2010 and again come up against the dominance of grey literature and how these initiatives are reported.

But the key finding from that work was widespread weaknesses and how initiatives are reported – weaknesses in terms of how they are designed. So very few programmes are explicitly based on clinical evidence; no programmes we assessed as partners mentioned reducing inequality as an explicit goal. And this is obviously a big issue in maternity services at the moment.

And I think it's an example of where there are gaps between what I describe as stated policy objectives in England and how programmes are actually designed.

So you have policy objectives around reducing inequality, involving in this case women and birth partners, patients and patient and public involvement, basing programmes on evidence.

But actually when you look in detail at the initiatives themselves, these sort of criteria for quality are met very variably. So I think there's an issue with how these large-scale change programmes in certainly maternity services and likely in other specialties as well are applied.

I think the next point I would make is probably around evaluation. It's building on our earlier discussion. And a big problem you have in maternity services is often a lot of these initiatives that attract resources in order to roll them out at scale often aren't evaluated. Actually, this includes a number of major national programmes.

That's an accountability problem in terms of how resources are used, but it's also a learning problem.

So again, coming back to what we were saying before, I think closing those gaps between how policymakers, practitioners and researchers conceptualise evidence, so what counts as evidence, what is around the threshold, what does that mean for evaluation? I think closing those gaps is really important, both from having an agreed basis in which we're learning from some of these programmes.  So those are my initial thoughts Graham, but happy to take follow up questions. I think Jane wants to come in here.

Jane: I think a lot of the big initiatives that you've described are coming from identification of , failure. So the genesis of this stuff is from identification of failure.

But then you have this slightly strange paradoxical juxtaposition which then when you do work to improve what you think, or rather improve on the outcome that you're trying to improve. So reducing inequalities or whatever it is, these are all laudable things, but we don't really necessarily understand why the system is achieving those things.

Often what happens when you learn, or you seek learn just from outcomes. You don't really know how everyday work is achieved. You don't know how the system every day is leading to these problems.

And so when you use a ‘find and fix’ approach, which is outcome based, you look into the system only when you have an outcome of interest or is deemed to be negative, then you might come up with different versions of the problem or indeed how the system is working than you would if you just went in and said, okay, so how do maternity services deliver their care every day? Like, what happens?

But you do get this then strange paradox, as you say, which is that when we've learned from failure, we don't then, and this linked to your work, Helen, we don't learn from what happens when you've actually tried to address that failure. So we're shaping what we look at and how we improve based on these negative outcomes.

But actually then we have this strange toxicity bias about, like, actually we're not interested necessarily in the evaluation of whether it's worked or not, because the process of doing something about it is much more important.

The visibility of that is much more important than the actual understanding necessarily of whether something has or hasn't worked. So it's, in a sense, quite a strange combination of things, where you have within healthcare – and rightly so, to some extent – a preoccupation with failure, and everything is really largely judged by the extent to which we have these failure events or not. But actually when we try and improve those failure events, we strangely don't really look at failure.

Graham: It's almost as if we need to turn the lens of improvement onto improvement itself. And I think all of us might know one or two academics who have said something along those lines.

To finish off, I wondered if I could ask each of you just very quickly, in the spirit of improving improvement, to suggest one thing that you think could improve the way in which we as researchers, or the system more broadly, draws on learning. Whether it's learning from success, learning from failure in a way that's really useful and can have an impact in practice. Helen, let me come to you first.

Helen: So one of the main preoccupations of my research is, is it possible to find a way to make doing this more easy for people so that we can spot things that go wrong often in quality improvement projects and diagnose issues and address issues on that basis in a more accessible way, thinking from the position that people, as I think Jane alluded to, don't even have time to get to the S in PDSA, quite often.

So they're doing something, checking to see if it works, and then basically that's it sometimes. So how can we help people to build in shortcuts to use this information, even if it's at quite a high level and we aren't going to ultimately produce some very detailed journal article about it?

Is there a way that we can help managers, clinicians, practitioners…to extract a nugget of valuable intelligence from something that didn't work and play that back in? So certainly that's a challenge that I'm setting myself, and I guess it remains to be seen whether I can come up with anything that gets anywhere near that.

Graham: But if you're taking us a step in that right direction or two steps in that right direction, then that's great. And I'm sure you'll make missteps and failures on the way and I'm sure the rest of the community will then step up and learn from them. It's all in that spirit of failing better, isn't it? So, James, what would be your one or two suggestions about how we can learn better as a system?

James: Learning better as a system? I think from my experience of the research I've been leading, I think it does come down to evaluation and sort of the motivation by my question earlier around closing these gaps and expectations between practitioners and researchers. I think too often, from what I've seen is certainly in maternity services, there's a disconnect between the resources that are allocated to improvement efforts, particularly at national level, and the emphasis placed, shall we say, on learning from those initiatives formally through evaluation.

And again, I think certainly at policy level or national level there's a recognition of the importance of valuation, but too often, in my view, that doesn't translate into to resources that are...that are allocated to it.

I think at service level that's often very frustrating because I think more often than clinicians who are engaged in QI day-to-day and services are well aware of the importance of accumulating learning from improvement efforts in a systematic fashion, but are too often under resourced to do so because for reasons we've discussed, they're trying to do this alongside, let's say, a clinical job and everything else they've got to do.

So I would say in terms of the main thing that the system, and I guess you have to think about the NHS and this context, needs to do to accumulate learning better is to ensure that resource allocation for learning or evaluation is proportionate to resources that have been spent on improvement efforts themselves.

Graham: Jane, final word to you.

Jane: I really like what James has just said. I possibly would add a couple more things. I think you know me, Graham, I'm a big fan of so-called Safety Two type approaches. I think we need in identifying problems to move away from simply looking at failure. I think we need to look at the identification of the everyday, the understanding of what happens every day.

I think the solutions that arise to improve services look very different potentially when you look at the everyday compared to when you look at failure. And then in terms of improvement, we need to flip it and say we need to be more preoccupied with failure than with necessarily toxic positivity.

So I think it's about better understanding and identification of problems through not just focusing on failure, but then focusing more on failure when we're actually seeking to understand what works.

And I think that links to what you're saying, James, around evaluation and having secured some funds for research to evaluate large scale policy initiatives in the last couple of years, the Patient Safety Incident Response Framework being one, I absolutely agree that so much work goes into shaping these policies and without a defined research effort, often these things might not be evaluated in quite the rigorous way that we want. But it's really important.

Graham: A big thank you to all the guests who joined today. Jane O'Hara, James McGowan and Helen Crump. I've been your host, Graham Martin and I hope you've enjoyed listening to Listen to THIS.

To find out more about THIS Institute's latest research, visit our website and head to the podcast page for show notes and to hear the series in full.

It's all thisinstitute.cam.ac.uk.