A library of logic models to explain how interventions to reduce diagnostic errors work

Maartje Kletter, MSc, University of Warwick

G.J. Mendelez-Torres, PhD, Cardiff University

Richard Lilford, PhD, University of Warwick

Celia Taylor, PhD, University of Warwick

Corresponding author: Celia Taylor, Warwick Medical School, University of Warwick, Coventry CV4 7AL.

Tel: 00442476524793, Email: celia.taylor@warwick.ac.uk

Acknowledgements

This research was funded by the NIHR Collaborations for Leadership in Applied Health Research and Care West Midlands initiative. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Abstract

Objectives: We aimed to create a library of logic models for interventions to reduce diagnostic error. This library can be used by those developing, implementing or evaluating an intervention to improve patient care, in order to understand what needs to happen, and in what order, if the intervention is to be effective.

Methods: To create the library we modified an existing method for generating logic models. Five ordered activities to include in each model were defined: pre-intervention, implementation of the intervention, post-implementation, but before the immediate outcome can occur, the immediate outcome (usually behaviour change) and post-immediate outcome, but before a reduction in diagnostic errors can occur. We also included reasons for lack of progress through the model. Relevant information was extracted about existing evaluations of interventions to reduce diagnostic error, identified by updating a previous systematic review.

Results: Data were synthesized to create logic models for four types of intervention, addressing five causes of diagnostic error in seven stages in the diagnostic pathway. In total 46 interventions from 43 studies were included and 24 different logic models were generated.

Conclusions: We used a novel approach to create a freely available library of logic models. The models highlight the importance of attending to what needs to occur before and after intervention delivery if the intervention is to be effective. Our work provides a useful starting point for intervention developers, helps evaluators identify intermediate outcomes and provides a method to enable others to generate libraries for interventions targeting other errors.

Key words: Diagnostic error, logic model, mechanistic theory, effectiveness

Word count: 3,981 (plus 1.044 in boxes)

Introduction

Any attempt to reduce the incidence of a particular error in healthcare must begin with an exploration of the epidemiology of the error, including an understanding of its cause, i.e. of *why* the particular error occurs [1]. It is then necessary to address the underlying cause by developing and implementing an appropriate *intervention* that changes the existing structure and/or process of care. In their review of methods for designing interventions intended to change the behaviour of healthcare professionals – the change required to address many (but not all) causes of error - Colquhoun and colleagues identified four tasks common to almost all methods: identification of barriers, selection of intervention components, use of theory and engagement of end-users [2]. These are time-consuming tasks. However, in many cases, an intervention developer does not have to start at square one because there are existing interventions that could be used (possibly following adaptation) for many error/cause of error combinations. To help a developer use an existing intervention with confidence, they need to know, amongst other things, how the intervention should be implemented, i.e. what specific steps are required and in what order, to make the intervention effective? This sequence of steps is known as the intervention's *logic model* or mechanistic theory [3-5]. In constructing a logic model, it is important to identify steps that need to occur *before* the intervention is implemented, as well as those that need to occur *after* the implementation if the final desired outcome is to be realised. A logic model should also include any specific facilitators and barriers that help or hinder progress at each step. By clearly specifying all of these steps, facilitators and barriers, logic models can also enable the identification of appropriate intermediate outcomes, such as fidelity, that should be measured during an evaluation to help explain the quantitative effect of the intervention on the final outcome (adverse events).

It has been argued that the use of logic models as part of theory-based intervention development will increase the probability that the intervention is effective [5, 6]. It is therefore good practice to describe an intervention's logic model in any report of its evaluation. However, including an explicit logic model is not prescribed in either the TIDieR [7] or the CONSORT [8] checklists.

The former stipulates that a full description of the intervention should be provided (including any essential theory), while the latter states that: "*Authors should … suggest a plausible explanation for how the intervention(s) might work, if this is not obvious*". Even a study adhering to both may result in the omission of important behavioural requirements, such as professionals' willingness to engage with the intervention. Therefore, although reports of evaluations of many existing interventions to reduce error are widely available, logic models are rarely included [9]. This lack of readily-accessible information makes it challenging for someone tasked with reducing a particular error to use an "off the shelf" intervention with confidence, just as it is challenging to bake a cake without a list of ingredients and recipe.

There are a number of systematic reviews that have considered the effectiveness of different possible interventions that aim to address specific types of error (see, for example, McDonald et al. on diagnostic errors [10], Royal et al. on prescribing errors in primary care [11] or Cottrell on wrong blood in tube errors in transfusion [12]). Although there are a number of patient safety practices with a strong evidence base [13], such practices do not yet exist for all errors. McDonald et al., for example, report that: "*some* interventions, …, *can* reduce diagnostic errors in *certain* situations" ([10], p. 382, emphasis added). Our premise is that one reason for the ineffectiveness of some interventions is that there is often insufficient attention afforded to the *full* logic model of the intervention i.e. from the decision to design and implement an intervention right through to a reduction in error at patient level [1, 6, 14, 15]. For example, while an effective training programme may have been developed, the intervention developers do not consider how to ensure all clinicians attend the training and subsequently apply their new knowledge once they are back in practice. We therefore aimed to show how full logic models for a range of existing interventions could be developed and compiled in a library, helping to broaden attention from intervention implementation alone to the entire intervention pathway. To illustrate our approach, we consider existing interventions that aim to address the causes of one specific error in healthcare, diagnostic error. We selected diagnostic errors because these are fairly common [16] and tend to have serious consequences [16-18]. Diagnostic errors have also been prioritised as a key focus for primary care by the WHO [19]. Our library can be used by intervention developers familiar with the specific type of diagnostic error they are aiming to address and its cause(s), to help them choose, modify and implement an appropriate intervention that addresses the cause of the error. By identifying the individual steps, the models should also "nudge" developers to ensure they can provide a sufficient justification (or causal theory) as to why each step in the model will lead to the next. The models in the library could also be used by intervention evaluators who need to know which intermediate outcome variables need to be measured. Our method for developing the models and synthesising them into a library can subsequently be used by other researchers seeking to create libraries of logic models of interventions addressing other types of error.

Methods

Search strategy for existing interventions to reduce diagnostic error

Our starting point was McDonald et al.'s systematic review of evaluations of interventions to reduce diagnostic error [10], which included 109 studies. This review only contained studies published before October 2012 and excluded studies in simulated settings. We therefore repeated the original search, and extended it to July 2016.

All of the titles and abstracts of the studies identified in our search were independently screened against a set of selection criteria (Box 1) by MK and CT. We used the inclusion criteria of McDonald et al., adapted to incorporate simulation-based studies, and added additional exclusion criteria designed to ensure the interventions included could be used in another setting (i.e. were not over-specific) and had data on their effectiveness available. We also excluded studies which increased the number of clinicians making an interpretation or changed the type of professional making the diagnosis, because of the minimal change to the diagnostic pathway that would result from implementing these interventions. The full text of all studies included by either reviewer was obtained and independently screened against the selection criteria by MK and CT.

Any disagreements regarding inclusion at the full text-stage were resolved by discussion and the reason for exclusion after full text screening was recorded.

Box 1: Selection criteria

<u>Inclusion criteria specified by McDonald et al. [10]</u>

Study evaluating any intervention to decrease diagnostic errors, the time to correct diagnosis or to appropriate clinical action.

Study in any clinical setting.

Any study design.

Study addressing patient-related outcomes or proxy measures of patient-related outcomes.

<u>Exclusion criteria specified by McDonald et al. [10]</u>

No intervention.

No real patients: modified for this review to include studies in simulated clinical settings and those with healthcare students as participants.

<u>Additional exclusion criteria for this review</u>

Studies where the intervention is a specific test used for a specific diagnosis.

Studies of interventions which increased the number of clinicians making an interpretation or changed the type of professional making the diagnosis.

Studies of evaluations of response to treatment or the effect of taking action on signs of deterioration.

Studies in which the intervention was designed primarily to reduce costs.

Studies not including an evaluation of the intervention.

Systematic (or other) reviews, case reports, letters, editorials, commentaries, opinion pieces, audits or protocols.

Generic library structure

In designing the structure of the library we considered the following course of action: a particular diagnostic error is identified, which could be due to one or more potential causes, each of which could be addressed with a number of potential interventions. The first level of the library therefore needed to describe the error itself, the second level the potential cause(s) of each error and the third level the types of intervention that could be implemented (Figure 1). Each logic model would then synthesize all of the specific interventions, of each type, that addressed each cause of each error. In order to operationalise this, we needed to create appropriate categories of errors (level 1), causes (level 2), and intervention types (level 3). For errors (level 1), we used the seven temporal stages (and sub-stages) of the diagnostic pathway as outlined by Schiff and colleagues [20]. For causes (level 2), we used an expanded version of the three-level categorisation outlined by Gandhi et al. [21] and Singh et al. [22] (cognitive, system-related and patient-related). We split cognitive causes into two categories, cognitive reasoning (akin to "judgment" in Gandhi et al.) and lack of knowledge/skill/experience ("lack of knowledge" in Gandhi et al.) because of the large number of interventions aiming to address cognitive-related errors. Furthermore, enhancing cognitive reasoning requires a different type of intervention to enhancing knowledge/ skill/experience. We added sub-optimal attention as a separate category, although we acknowledge that this may not accord with "no blame" patient safety cultures. This provided five "error cause" categories in total. For intervention type (level 3), we used a modified version of the six categories outlined by McDonald et al. [10]. The educational and technology intervention categories were retained unchanged. We amalgamated personnel and technique changes into the process change category and added quality improvement interventions as a separate category. Studies using only additional review methods were excluded (as discussed above) to give four "intervention type" categories in total.

The seven diagnostic pathway stages, five causes of error and four types of intervention meant that our library could theoretically contain up to 7x5x4 = 140 logic models..

CT and MK subsequently independently coded each intervention using these three categorisations; each intervention in studies including multiple interventions was coded separately. Results were then compared and any disagreements resolved by discussion. Information on the following additional aspects of each intervention was coded by MK, using

NVivo Pro v11: specific intervention description, setting (including whether a simulation), participants and study design. In addition MK coded any *ex ante* explanation of why the intervention was expected to work and any *ex post* explanation of why the intervention did or did not work. All coding was subsequently verified by CT.

Logic model structure and generation of synthesized logic models

We applied a modified version of Kneale et al.'s procedure for logic model creation [9], as described in Box 2, with the aim of identifying, in the most plausible temporal order, the actitvities that would be included in intervention development and implementation. We decided that the starting point for each model would be the *decision* to implement a specific intervention and subsequently identified five key temporal activities to include in each model: pre-intervention (intervention development and other requirements before the intervention can be implemented on the ground), the implementation of the intervention itself, post-implementation (what needs to happen before the immediate outcome of the intervention can occur), the immediate outcome (which generally mitigated the underlying cause of the error) and post-immediate outcome (before the effects can reach the patient and a reduction in diagnostic errors can occur). Within each stage, there could be multiple steps (i.e. the individual requirements, activities and/or changes). This meant that each logic model would show the full, ordered chain by which intervention implementation leads to the desired outcome.

We modified Kneale et al.'s procedures in three ways. First, following examination of logic models in existing studies and general frameworks (#1 in Box 2), we worked forwards from the initial design of the intervention to the final (distal) outcome, rather than the other way round, as this seems a better match to what an implementer would do in practice having chosen a specific intervention. Second, we extended #8 (sharing initial logic models) to include the generation of a single, synthesized model for each error/cause/type of intervention combination. Finally, we excluded #10 (presenting the final logic model in the protocol for the review) as it was not required for our work. We also wanted to include an indication of the effectiveness of each intervention, to aid users of the library in selecting a potentially effective intervention. Our method of doing so is described in Box 3.

Box 2: Generation of synthesized logic models

#1: Examination of logic models in existing studies and general frameworks: We gathered the coded explanations for intervention (in)effectiveness from our NVivo database. Given that the majority of interventions sought to achieve some form of professional behaviour change, we also examined the COM-B framework [23], the Stages of Change model for behavioural change interventions [24] and Kirkpatrick's hierarchy of outcomes for educational interventions [25]. These explanation, frameworks and models provided an overview of the individual steps that needed to be included in our logic models in each of the five key activities we had already identified.

[For #2 to #5, CT and MK worked independently, aggregating the information from #1 to enable development of a draft logic model for each intervention in each study.]

#2: Specification of intervention inputs (intervention development and other requirements before the intervention can be implemented on the ground): We identified two main types of input: suitable intervention design and the intended subjects being able to attend to it. Drawing on the COM-B framework [23] for example, the curriculum and pedagogy of a training programme (as an example of a specific intervention) would need to be appropriate to enable the development of the psychological capacity of the target audience and the intended "subjects" of the intervention would need sufficient time (social opportunity) to attend to it.

#3: Specification of intervention processes: This is an explanation of how the intervention would be provided (e.g. the nature of the training provided to clinicians) and what resources would be required in order to do so (e.g. room space).

#4: Identification of what needs to happen post-implementation, before the immediate outcome of the intervention can occur: We identified any requirements for those using the intervention in practice,

including Kneale et al.'s "proximal" outcomes [9]. Drawing on the Kirkpatrick model for training evaluation [25], our exemplar training programme could only be effective if clinicians were engaged during the course and learnt from it.

#5: Identification of immediate outcome and steps from the immediate to the distal outcome: Our "immediate" outcome was equivalent to Kneale et al.'s "intermediate" outcome [9], the change necessary to achieve the distal (final) outcome (usually behaviour change). Such behaviour change is the "action" stage in the stages of change model [24], the third level in the Kirkpatrick model [25] and the outcome of the COM-B framework [23].

#6: Identification of distal outcome: We had already identified a common distal outcome for all interventions, a reduction in diagnostic errors impacting on patient-level outcomes. This would be achieved when a clinician made a correct or timelier diagnosis that they would not have done in the absence of the intervention.

#7: Specification of intervention moderators including setting and population group: To avoid over-complication, we did not include these aspects within the logic models themselves but extracted information on setting and participants, as described above and which are presented separately.

#8: Share initial logic models, review and generate a single, synthesized model for each error/cause/type of intervention combination: MK and CT shared the logic models they had developed for each intervention and discussed similarities and differences. We then agreed on a model for each error/cause/type of intervention combination as shown in Figure 1. Within the "testing" error category we developed one logic model for each sub-category to avoid over-complication.

#9: Share synthesized models with the whole group, review and revise: The synthesized models were then shared with the remainder of the team and revised as required.

Box 3: Determining intervention effectiveness

The effectiveness of the interventions was assessed based on the size of the effect achieved and its statistical significance. For an intervention's effect size (ES), we used results for total diagnostic accuracy or for all errors combined (including all 'levels' of error from minor to major) and across all participants (rather than for a specific type of error or a specific participant sub-group), unless there was a clear indication in the study that the primary outcome was for a specific type of error/sub-group. If studies included immediate and longitudinal effects, we used outcomes measured immediately after the intervention, as not all studies included repeat measurements and the time gaps where this was done were variable. The outcome we used (detailed in Appendix 1) was not always that reported in the abstract of the paper. For some papers we used the primary data presented to calculate effect size and statistical significance, using the Campbell Collaboration's effect size calculator, using the logit method for 2x2 tables and pooled standard deviations for paired t-tests [26]. Any effective intervention was shown as having a positive effect size, regardless of whether the outcome related to diagnostic accuracy or error rates. It was not always possible to determine effect size and statistical significance from the results or data presented and in some cases we were unable to adjust for non-independence in pre/post studies where the same participants contributed data in both time periods, albeit regarding different (simulated) patients. Using Cohen's rules of thumb [27] and traditional frequentist approaches to determining statistical significance, we classified the effectiveness of the intervention as negative ($ES<0$ and $p<0.05$), none ($p>0.05$), very small ($0<ES<0.2$ and $p<0.05$), small ($0.2<ES<0.5$ and $p<0.05$), medium ($0.5<ES<0.8$ and $p<0.05$) or large ($ES>0.8$ and $p<0.05$).

Results

We reviewed 2,638 titles and abstracts and 286 full text studies. A total of 43 studies met the inclusion criteria (Figure 2) and proceeded to data extraction and coding. Of the 140 potential logic models, there was at least one intervention in 19 (14%). A total of 58 active trial arms were reported across the 43 studies. After grouping very similar interventions, a total of 46 unique (specific) interventions were identified.

Table 1 summarises the studies included in the logic models in each combination; full details on each are provided in Appendix 1. The most common errors addressed were errors in the testing stage of the diagnostic pathway (N=26 interventions, 60%). The most common interventions addressed errors caused by a lack of knowledge/skill/experience (N=18, 39%) or sub-optimal cognitive reasoning (N=14, 30%). The most common types of interventions were those in the process category (N=18, 39%) and the education and feedback category (N=16, 35%).

51 effect sizes could be calculated although some were for multi-component interventions as a whole. While no interventions had a statistically significant negative effect, only seven (14%) were classified as having "large" effect sizes and 16 (31%) were classified as having no effect.

An example of a logic model, for errors in diagnostic decision making caused by sub-optimal cognitive reasoning and addressed with education and feedback interventions, is shown in Figure 3. The full library of the 24 generated logic models is shown in Appendix 2. All logic models use the generic term "clinician" to denote any healthcare professional or staff member involved in making a diagnosis at any stage in the diagnostic pathway. To generate the logic model shown in Figure 3, we drew on two specific interventions in this error-cause-type combination, a training programme in diagnostic coding for psychiatric disorders (ICD-10) trialled in a simulated setting [28] and cognitive forcing strategy training trialled with medical students in a simulated emergency medicine setting [29]. The use of a structured diagnostic system (i.e. ICD-10 codes) was intended to help overcome the cultural biases known to affect diagnostic decision-making in psychiatry [28]. The cognitive forcing training aimed to encourage participants to use analytic, or System 2, thinking during diagnostic reasoning, which means that they would self-monitor following an initial diagnosis and "force" themselves to consider any alternative, non-obvious diagnoses [29]. At the pre-intervention stage each training programme needed to be designed appropriately in terms of curriculum and pedagogy and participants needed to be given time to attend the training. During the intervention stage training would be provided. Clinicians needed to actually attend the training, engage in it (e.g. pay attention), learn from the training and retain this learning. The immediate outcome would be that the participants change their existing behaviour by applying the newly learnt knowledge/skills in diagnostic decision making. During the post-immediate outcome stage the use of the learnt knowledge/skills would need to help the clinician make a correct diagnosis (that they would not have done previously), if the intervention is to reduce diagnostic error.

The effectiveness of both specific interventions included in Figure 3 was evaluated in simulations of clinical practice using a test requiring participants to diagnose one or more cases, with one showing a large effect [28] and the other no effect [29]. Sherbino and colleagues [29] suggested a number of reasons why their intervention was ineffective, including insufficiently complex cases that did not require System 2 thinking, a lack of transfer of learning to new cases and an insufficiently strong training programme to counter existing cognitive biases. For the intervention found to be effective [28], it would still be necessary to show longer-term retention and transfer to real-life clinical practice if patient-level outcomes are to be improved.

Discussion

Summary of findings

We have generated 24 logic models which show the mechanistic theory of 46 different interventions designed to reduce the incidence of diagnostic error in healthcare. These models can be used by anyone seeking to develop and implement an intervention to reduce a specific diagnostic error in their own setting. The models provide a guide as to what needs to be done in what order if the desired final effects of a particular intervention are to be realised; as such they also help intervention evaluators choose appropriate intermediate outcomes. One prerequisite for using the library is that the intervention developer has a good idea of the main cause of the error they are trying to tackle; although of course many errors are multi-factorial [30]. Intervention developers also need to be cognisant of how any aspects of their own context may mean that the intervention has a different level of effectiveness to that in the evaluations included in this study. Thus, while a developer may need to adapt an existing intervention, they do not have to start with a blank piece of paper.

As with patient safety incidents, which are often followed-up with investigations using techniques such as Root Cause Analysis [31], we can learn from the unsuccessful interventions by examining the "leaks" from the logic models. For example, in Goodacre et al.'s study [32], computer-generated interpretations of ECG results were provided to clinicians but one reason for a lack of intervention effectiveness was that the results were ignored. In general, however, there was a lack of evidence in the included studies about potential "leaks", as has also been noted by others [33]. An intervention developer wanting to implement a similar intervention in their own context should therefore be encouraged to discuss the proposed ECG reports with clinicians and determine whether they would be used and why/why not; and to consider any other leaks that may occur at other steps in the logic model. The final intervention design and implementation would also need to include a strategy to improve adherence, such as routine reminders or peer assistance.

Strengths

To our knowledge, this is the first attempt at creating and providing a library of logic models which enables a user to compare and contrast different interventions and to understand what needs to occur and in what order if an intervention is to be effective. Our task was more challenging than we had originally anticipated, as none of the included studies explicitly described the full logic model for the intervention being evaluated. By using the library, intervention developers should be able to develop and implement interventions that are more likely to be effective, as they can ensure that all steps in the logic model are considered at an early stage.

Limitations

We were only able to generate 24 logic models. There will be more potential models, because interventions for other meaningful error/cause combinations are yet to be developed and/or evaluated. The existing breakdown of interventions by type of diagnostic error may not match the prevalence or severity of different types of error in reality. The library should therefore be updated when evidence accumulates, although some of the cells in Table 1 may be empty because a particular error is unlikely to be due to a particular cause (e.g. missing information on samples is unlikely to be due to cognitive bias because the cognitive load of completing the information required is low). Nevertheless, the "gaps" in Table 1 could be combined with evidence on the epidemiology of error to identify priorities for intervention development. Although we followed a standardized procedure for generating the logic models, and based our model structure on existing work [10, 20-22], they remain subjective and could be challenged by others. In particular, many errors have multiple causes (as identified by Graber et al. [30]) but we assigned each intervention to only one overall cause category. However, some interventions address more than one possible cause of each error and we would encourage intervention developers to consider all possible causes and design multi-faceted interventions when required.

We also advocate greater adherence to the TIDieR checklist [7], as clearer intervention descriptions would have enabled us to provide more objective logic models.

We have not included causal theories in our logic models, as we discuss in more detail below. Our approach suggests that intervention implementation through the steps in the logic model is linear in time, when this is unlikely to be the case for all interventions in practice. Although we provided an indication of each study's effectiveness, it was outside our remit to determine which specific components of multi-faceted interventions were critical for overall effectiveness, however it is also plausible that the "effectiveness sum" of a multi-faceted intervention is greater than that of the sum of its parts and, indeed, multi-faceted interventions may well be essential [34]. Likewise, we do not yet know the relative importance of each step in a logic model or the impact of context on effectiveness; other authors have reported a paucity of evidence in this area across patient safety interventions more generally [33]. Furthermore, we did not undertake a quality appraisal of the included studies, so our estimates of effectiveness may be biased.

The sample of studies (and therefore interventions) included was limited by our inclusion criteria; for example we excluded studies of interventions that focused on reducing costs without increasing the error rate or in which the only intervention was to increase the number of clinicians reviewing test results prior to making a diagnosis. Our sample may also be limited by publication bias, which is likely to reduce the number of ineffective interventions included. While a user of the library may be less likely to choose an intervention previously found to be ineffective, their inclusion would help us to learn from previous mistakes.

Comparison with existing literature and future work

It is generally accepted that all interventions should be based on causal theory [6, 10, 14, 15], and knowing an intervention's logic model or mechanistic theory is a prerequisite for explaining its causal theory (i.e. we need to identify the steps in the logic model before we can explain the "why" of each; bearing in mind that different causal theories may be needed to link different pairs of steps). However, the superior effectiveness of theory-led over non-theory-led interventions is not always borne out in practice [3]. Our work suggests that one reason for this is that while a theory-based intervention may make the "immediate" outcome of the intervention more likely (e.g. the knowledge level of the clinicians who attend an educational intervention increases), there are additional steps both before and after the intervention itself where various "leaks" from the logic model dilute effectiveness.

There are four possible extensions to the work presented here. The first is to apply our method to interventions designed to tackle different errors, such as prescribing errors, and subsequently, to synthesise results across these different errors in the context of patient safety in general. The second is to identify which steps in the logic model, context and intervention design features are critical for effectiveness, and which tend to lead to ineffectiveness, potentially using Qualitative Comparative Analysis [35]. This task will however be difficult given the large variety of interventions and types of error across the included studies. Third, we could identify plausible causal theories for each link in each logic model. Again this will not be a simple task; Michie and colleagues, for example have identified and described 83 theories of behaviour change [36]. Finally, we could consider the quantitative relationships between steps in the logic models. For example, the logic models could be presented as Bayesian networks, which would facilitate the synthesis of multiple sources of evidence to derive estimates of the effect on the intervention on health outcomes and costs [37].

Conclusion

We were able to generate logic models for all of the interventions to reduce diagnostic error identified in our search and the resulting library is freely available to all (Appendix 2). We had to rely on the published evaluation reports for information about each intervention, meaning that logic model development was partially subjective. However, we based our method on previously published work [9], although we worked in the opposite direction to Kneale and colleagues, from intervention design to distal outcome. The resulting library of logic models can be used by others in a variety of ways: the library gives intervention developers a useful starting point and encourages them to consider and publish their logic models and identify appropriate causal theories, and helps intervention evaluators to identify and measure critical intermediate outcome measures. Furthermore the methods we have described will help researchers to generate libraries for interventions targeting other errors in healthcare.

Figure legends

Figure 1: Generic library structure

Figure 2: Flow diagram

Figure 3: Logic model for errors in diagnostic decision making caused by cognitive bias and addressed with education and feedback interventions [28, 29]

References

1. Brown, C., et al., *An epistemology of patient safety research: a framework for study design and interpretation. Part 1. Conceptualising and developing interventions.* Qual Saf Health Care, 2008. **17**.
2. Colquhoun, H.L., et al., *Methods for designing interventions to change healthcare professionals' behaviour: a systematic review.* Implementation Science, 2017. **12**(1): p. 30.
3. Brown, H.E., et al., *Family-based interventions to increase physical activity in children: a systematic review, meta-analysis and realist synthesis.* obesity reviews, 2016. **17**(4): p. 345-360.
4. Anderson, L.M., et al., *Using logic models to capture complexity in systematic reviews.* Research synthesis methods, 2011. **2**(1): p. 33-42.
5. Bruce, B.B., et al., *Methodologies for evaluating strategies to reduce diagnostic error: report from the research summit at the 7th International Diagnostic Error in Medicine Conference.* Diagnosis, 2016. **3**(1): p. 1-7.
6. Foy, R., et al., *The role of theory in research to develop and evaluate the implementation of patient safety practices.* BMJ Quality & Safety, 2011. **20**(5): p. 453-459.
7. Hoffmann, T.C., et al., *Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide.* Bmj, 2014. **348**: p. g1687.
8. Moher, D., et al., *CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials.* Journal of clinical epidemiology, 2010. **63**(8): p. e1-e37.
9. Kneale, D., J. Thomas, and K. Harris, *Developing and Optimising the Use of Logic Models in Systematic Reviews: Exploring Practice and Good Practice in the Use of Programme Theory in Reviews.* PloS one, 2015. **10**(11): p. e0142187.
10. McDonald, K.M., et al., *Patient safety strategies targeted at diagnostic errors: a systematic review.* Annals of internal medicine, 2013. **158**(5_Part_2): p. 381-389.
11. Royal, S., et al., *Interventions in primary care to reduce medication related adverse events and hospital admissions: systematic review and meta-analysis.* Quality and Safety in Health Care, 2006. **15**(1): p. 23-31.

12. Cottrell, S., et al., *Interventions to reduce wrong blood in tube errors in transfusion: a systematic review.* Transfusion medicine reviews, 2013. **27**(4): p. 197-205.
13. Shekelle, P.G., et al., *The top patient safety strategies that can be encouraged for adoption now.* Annals of Internal Medicine, 2013. **158**(5_Part_2): p. 365-368.
14. Reed, J.E., et al., *Designing quality improvement initiatives: the action effect method, a structured approach to identifying and articulating programme theory.* BMJ quality & safety, 2014. **23**(12): p. 1040-1048.
15. Medical Research Council, *Developing and evaluating complex interventions: new guidance*. 2008, Medical Research Council: London.
16. Zwaan, L., et al., *Patient record review of the incidence, consequences, and causes of diagnostic adverse events.* Archives of Internal Medicine, 2010. **170**(12): p. 1015-1021.
17. Tehrani, A.S.S., et al., *25-Year summary of US malpractice claims for diagnostic errors 1986–2010: an analysis from the National Practitioner Data Bank.* BMJ Quality & Safety, 2013. **22**(8): p. 672-680.
18. Panesar, S.S., et al., *How safe is primary care? A systematic review.* BMJ Quality & Safety, 2016. **25**(7): p. 544-553.
19. Cresswell, K.M., et al., *Global research priorities to better understand the burden of iatrogenic harm in primary care: an international Delphi exercise.* PLoS medicine, 2013. **10**(11): p. e1001554.
20. Schiff, G.D., et al., *Diagnosing diagnosis errors: lessons from a multi-institutional collaborative project.* 2005.
21. Gandhi, T.K., et al., *Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims.* Annals of internal medicine, 2006. **145**(7): p. 488-496.
22. Singh, H., et al., *System-related interventions to reduce diagnostic errors: a narrative review.* BMJ Qual Saf, 2011: p. bmjqs-2011-000150.

23. Michie, S., M.M. van Stralen, and R. West, *The behaviour change wheel: a new method for characterising and designing behaviour change interventions.* Implementation Science, 2011. **6**(1): p. 42.

24. Norcross, J.C., P.M. Krebs, and J.O. Prochaska, *Stages of change.* Journal of clinical psychology, 2011. **67**(2): p. 143-154.

25. Hammick, M., T. Dornan, and Y. Steinert, *Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No 13*. ND, BEME Collaboration.

26. Wilson, D. *Practical Meta-Analysis Effect Size Calculator [Online calculator].* . ND 03/03/2017]; Available from: https://www.campbellcollaboration.org/this-is-a-web-based-effect-size-calculator/explore/this-is-a-web-based-effect-size-calculator.

27. Cohen, J., *Statistical power analysis for the behavioral sciences*. 1988, Hillsdale, NJ: Lawrence Erlbaum.

28. Rezvyy, G., et al., *Correcting biases in psychiatric diagnostic practice in Northwest Russia: Comparing the impact of a general educational program and a specific diagnostic training program.* BMC medical education, 2008. **8**(1): p. 1.

29. Sherbino, J., et al., *Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial.* CJEM, 2014. **16**(01): p. 34-40.

30. Graber, M.L., N. Franklin, and R. Gordon, *Diagnostic error in internal medicine.* Archives of internal medicine, 2005. **165**(13): p. 1493-1499.

31. Wilson, P.F., *Root cause analysis: A tool for total quality management*. 1993: ASQ Quality Press.

32. Goodacre, S., A. Webster, and F. Morris, *Do computer generated ECG reports improve interpretation by accident and emergency senior house officers?* Postgraduate medical journal, 2001. **77**(909): p. 455-457.

33. Øvretveit, J.C., et al., *How does context affect interventions to improve patient safety? An assessment of evidence from studies of five patient safety practices and proposals for research.* Quality and Safety in Health Care, 2011. **20**(7): p. 604-610.

34. Singh, H., et al., *The global burden of diagnostic errors in primary care.* BMJ Qual Saf, 2016: p. bmjqs-2016.

35. Thomas, J., A. O'Mara-Eves, and G. Brunton, *Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example.* Systematic reviews, 2014. **3**(1): p. 67.

36. Michie, S., et al., *ABC of behaviour change theories*. 2014, Sutton: Silverback Publishing.

37. Watson, S.I. and R.J. Lilford, *Integrating multiple sources of evidence: a Bayesian perspective*, in *Challenges, solutions and future directions in the evaluation of service innovations in health care and public health*, R. Raine, R. Fitzpatrick, and H.e.a. Barratt, Editors. 2016, NIHR: Southampton.

Table 1: Summary of error, cause of error and intervention types

		Cause of error				
Stage in diagnostic process (Schiff)	**Error (Schiff sub-category; only sub-categories with at least one intervention are included)**	**Sub-optimal cognitive reasoning**	**Lack of knowledge/ skill/ experience**	**Sub-optimal attention**	**System-related**	**Patient-related**
Access/ presentation	N/A					
History taking	Failure/delay in eliciting critical piece of history data					**P**: Patient-completed questionnaire [1-3]
Physical exam	Failure/delay in eliciting critical physical exam finding			**EF**: Patient feedback [4]		
	Sub-optimal weighting				**P**: Tertiary trauma survey [5, 6]	
Testing	Failure/delay in performing ordered tests			**T**: Computer test support [7]		
	Sample mix-up/mislabelled			**P**: Computer-aided double-signing [8] **T**: Computer test support [9]		
	Technical errors/poor processing of specimen/test		**EF**: Poster with most common errors [10]; Crash course about most common errors [10]; Leaflet explaining blood drawing procedure and explanation of procedure by senior nurse [7]; Training on sample management and standardized sample collection [8]; Reference materials on sample collection produced [8]; Training on blood sample collection [11, 12]		**P**: Improved storage facilities [8]; More delivery staff [8] **QI**: Participation in cross-institution benchmarking [13]	
	Failed/delayed transmission of result to clinician			**P**: Structured report template [14]	**P**: Quiet working environment [14]	

	Erroneous clinician interpretation of test	**P**: Verification stage added [15]; Checklists to correct mistakes in initial diagnosis [15, 16] **T**: Computer pattern recognition [17]	**EF**: Individual feedback on image interpretation [18, 19]; Meetings to discuss errors/missed cases [20, 21]; Technician report written at time of investigation and presented to clinicians [22]; Training including hands-on training and expert tutorial [23] **T**: Software to help trainees read capsule endoscopy images [19]; Computer test support [24]; Computer-interpretation of investigation results provided to clinicians [25, 26]		**P**: Structured reporting process [20] **T**: Computerised version of images [27]	
Assessment	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	**EF**: Specific training programme in diagnostic coding [28]; Cognitive forcing strategy training [29] **P**: Self-directed reflection [30]; Enhanced analytical reasoning using structured template [31]; Provision of additional data and querying initial hypothesis [32]; Structured reanalysis of case findings [33]; Checklists after collecting information without return to patient [34]; Checklists after collecting information with return to patient [34] **T**: Diagnostic reminder system [35, 36]; Computer diagnostic support system before testing [37, 38]; Computer diagnostic support system after testing [37, 38]	**EF**: Monthly feedback added to standardised data collection and computer support [39]; Education about atypical presentations [40]; Feedback about telephone follow-up of high risk patients [40] **P**: Standardised data collection forms [39] **T**: Computer-based decision support tool [39, 41, 42]			
Referral	Failure/delay in ordering needed referral			**P**: Reminders [43]		
Follow-up	N/A					

Type of intervention codes: Education/feedback (EF), Process (P), Technology (T), Quality improvement activities (QI).
Several interventions were sufficiently similar in multiple studies to group them as one intervention and the number of references specifies the number of studies, including two sets of two papers [35-38] in which the interventions were identical. Some studies included multiple interventions (range 1-5). Where the second and any subsequent interventions built on the first, the intervention is coded according to its incremental type. N/A: No interventions in this stage of the diagnostic process identified.

References

1. Lewis, G., et al., *Computerized assessment of common mental disorders in primary care: effect on clinical outcome.* Family Practice, 1996. **13**(2): p. 120-126.
2. Mueller, C.A., et al., *Disclosure of new health problems and intervention planning using a geriatric assessment in a primary care setting.* Croatian medical journal, 2010. **51**(6): p. 493-500.
3. Schriger, D.L., et al., *Enabling the diagnosis of occult psychiatric illness in the emergency department: a randomized, controlled trial of the computerized, self-administered PRIME-MD diagnostic system.* Annals of emergency medicine, 2001. **37**(2): p. 132-140.
4. Nicholl, D., et al., *The TOS study: can we use our patients to help improve clinical assessment?* The journal of the Royal College of Physicians of Edinburgh, 2011. **42**(4): p. 306-310.
5. Biffl, W.L., D.T. Harrington, and W.G. Cioffi, *Implementation of a tertiary trauma survey decreases missed injuries.* Journal of Trauma and Acute Care Surgery, 2003. **54**(1): p. 38-44.
6. Keijzers, G.B., et al., *A prospective evaluation of missed injuries in trauma patients, before and after formalising the trauma tertiary survey.* World journal of surgery, 2014. **38**(1): p. 222-232.
7. Lillo, R., et al., *Reducing preanalytical laboratory sample errors through educational and technological interventions.* 2012.
8. Li, H.y., et al., *Reduction of preanalytical errors in clinical laboratory through multiple aspects and whole course intervention measures.* Journal of Evidence-Based Medicine, 2014. **7**(3): p. 172-177.
9. Turner, H.E., et al., *The effect of electronic ordering on pre-analytical errors in primary care.* Annals of Clinical Biochemistry: An international journal of biochemistry and laboratory medicine, 2013: p. 0004563213494184.
10. Hlabangana, L.T. and S. Andronikou, *Short-term impact of pictorial posters and a crash course on radiographic errors for improving the quality of paediatric chest radiographs in an unsupervised unit—a pilot study for quality-assurance outreach.* Pediatric radiology, 2015. **45**(2): p. 158-165.
11. Hopkins, K., et al., *Reducing blood culture contamination rates: a systematic approach to improving quality of care.* American journal of infection control, 2013. **41**(12): p. 1272-1274.
12. Ramirez, P., et al., *Blood culture contamination rate in an intensive care setting: Effectiveness of an education-based intervention.* American journal of infection control, 2015. **43**(8): p. 844-847.
13. Raab, S.S., et al., *The value of monitoring frozen section-permanent section correlation data over time.* Archives of pathology & laboratory medicine, 2006. **130**(3): p. 337-342.
14. Rosskopf, A.B., et al., *Quality Management in Musculoskeletal Imaging: Form, Content, and Diagnosis of Knee MRI Reports and Effectiveness of Three Different Quality Improvement Measures.* American Journal of Roentgenology, 2015. **204**(5): p. 1069-1074.
15. Sibbald, M., A.B. de Bruin, and J.J. van Merrienboer, *Checklists improve experts' diagnostic decisions.* Medical education, 2013. **47**(3): p. 301-308.
16. Sibbald, M., A.B. De Bruin, and J.J. van Merrienboer, *Finding and fixing mistakes: do checklists work for clinicians with different levels of experience?* Advances in Health Sciences Education, 2014. **19**(1): p. 43-51.
17. Kundel, H.L., C.F. Nodine, and E.A. Krupinski, *Computer-displayed eye position as a visual aid to pulmonary nodule interpretation.* Investigative radiology, 1990. **25**(8): p. 890-896.

18. Tudor, G.R. and D.B. Finlay, *Error review: can this improve reporting performance?* Clinical radiology, 2001. **56**(9): p. 751-754.
19. Hosoe, N., et al., *Evaluations of capsule endoscopy software in reducing the reading time and the rate of false negatives by inexperienced endoscopists.* Clinics and research in hepatology and gastroenterology, 2012. **36**(1): p. 66-71.
20. Espinosa, J.A. and T.W. Nolan, *Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study.* Bmj, 2000. **320**(7237): p. 737-740.
21. Itri, J.N., et al., *Using focused missed-case conferences to reduce discrepancies in musculoskeletal studies interpreted by residents on call.* American Journal of Roentgenology, 2011. **197**(4): p. W696-W705.
22. Dudley, M. and K. Channer, *Assessment of the value of technician reporting of electrocardiographs in an accident and emergency department.* Journal of accident & emergency medicine, 1997. **14**(5): p. 307-310.
23. Rondonotti, E., et al., *Can we improve the detection rate and interobserver agreement in capsule endoscopy?* Digestive and Liver Disease, 2012. **44**(12): p. 1006-1011.
24. Nishikawa, R.M., et al., *Clinically missed cancer: how effectively can radiologists use computer-aided detection?* American Journal of Roentgenology, 2012. **198**(3): p. 708-716.
25. Tsai, T.L., D.B. Fridsma, and G. Gatti, *Computer decision support as a source of interpretation error: the case of electrocardiograms.* Journal of the American Medical Informatics Association, 2003. **10**(5): p. 478-483.
26. Goodacre, S., A. Webster, and F. Morris, *Do computer generated ECG reports improve interpretation by accident and emergency senior house officers?* Postgraduate medical journal, 2001. **77**(909): p. 455-457.
27. Weatherburn, G., et al., *The effect of a picture archiving and communications system (PACS) on diagnostic performance in the accident and emergency department.* Journal of accident & emergency medicine, 2000. **17**(3): p. 180-184.
28. Rezvyy, G., et al., *Correcting biases in psychiatric diagnostic practice in Northwest Russia: Comparing the impact of a general educational program and a specific diagnostic training program.* BMC medical education, 2008. **8**(1): p. 1.
29. Sherbino, J., et al., *Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial.* CJEM, 2014. **16**(01): p. 34-40.
30. Monteiro, S.D., et al., *Reflecting on diagnostic errors: taking a second look is not enough.* Journal of general internal medicine, 2015. **30**(9): p. 1270-1274.
31. Myung, S.J., et al., *Effect of enhanced analytic reasoning on diagnostic accuracy: a randomized controlled study.* Medical teacher, 2013. **35**(3): p. 248-250.
32. Coderre, S., B. Wright, and K. McLaughlin, *To think is good: querying an initial hypothesis reduces diagnostic error in medical students.* Academic Medicine, 2010. **85**(7): p. 1125-1129.
33. Mamede, S., et al., *Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents.* Jama, 2010. **304**(11): p. 1198-1203.
34. Sibbald, M., et al., *Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam.* BMJ quality & safety, 2013. **22**(4): p. 333-338.

35. Ramnarayan, P., et al., *Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: a quasi-experimental study.* BMC Medical Informatics and Decision Making, 2006. **6**(1): p. 1.
36. Ramnarayan, P., et al., *Diagnostic omission errors in acute paediatric practice: impact of a reminder system on decision-making.* BMC medical informatics and decision making, 2006. **6**(1): p. 1.
37. Kostopoulou, O., et al., *Early diagnostic suggestions improve accuracy of family physicians: a randomized controlled trial in Greece.* Family practice, 2015: p. cmv012.
38. Kostopoulou, O., et al., *Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients.* Br J Gen Pract, 2015. **65**(630): p. e49-e54.
39. Wellwood, J., S. Johannessen, and D. Spiegelhalter, *How does computer-aided diagnosis improve the management of acute abdominal pain?* Annals of the Royal College of Surgeons of England, 1992. **74**(1): p. 40.
40. Chern, C.-H., et al., *Decreasing clinically significant adverse events using feedback to emergency physicians of telephone follow-up outcomes.* Annals of emergency medicine, 2005. **45**(1): p. 15-23.
41. Segal, M.M., et al., *Evidence-based decision support for neurological diagnosis reduces errors and unnecessary workup.* Journal of child neurology, 2014. **29**(4): p. 487-492.
42. Wexler, J.R., et al., *Impact of a system of computer-assisted diagnosis: initial evaluation of the hospitalized patient.* American Journal of Diseases of Children, 1975. **129**(2): p. 203-205.
43. Murphy, D.R., et al., *Electronic trigger-based intervention to reduce delays in diagnostic evaluation for cancer: a cluster randomized controlled trial.* Journal of Clinical Oncology, 2015. **33**(31): p. 3560-3567.

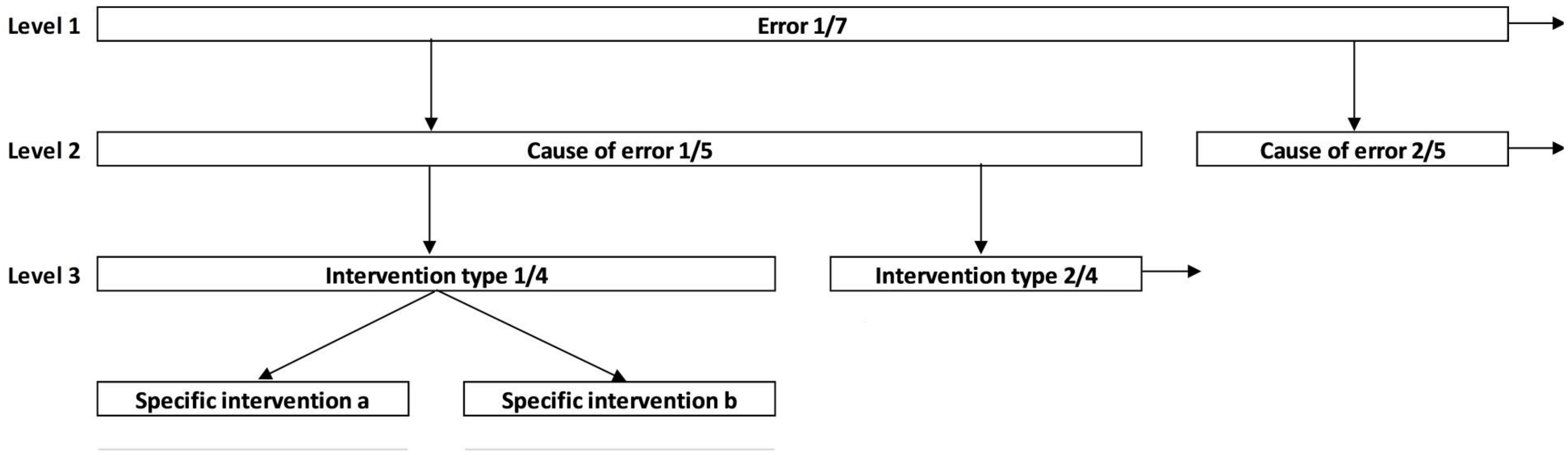

Level 1
Error 1/7
Level 2
Cause of error 1/5
Cause of error 2/5
Level 3
Intervention type 1/4
Intervention type 2/4
Specific intervention a
Specific intervention b

Figure 1

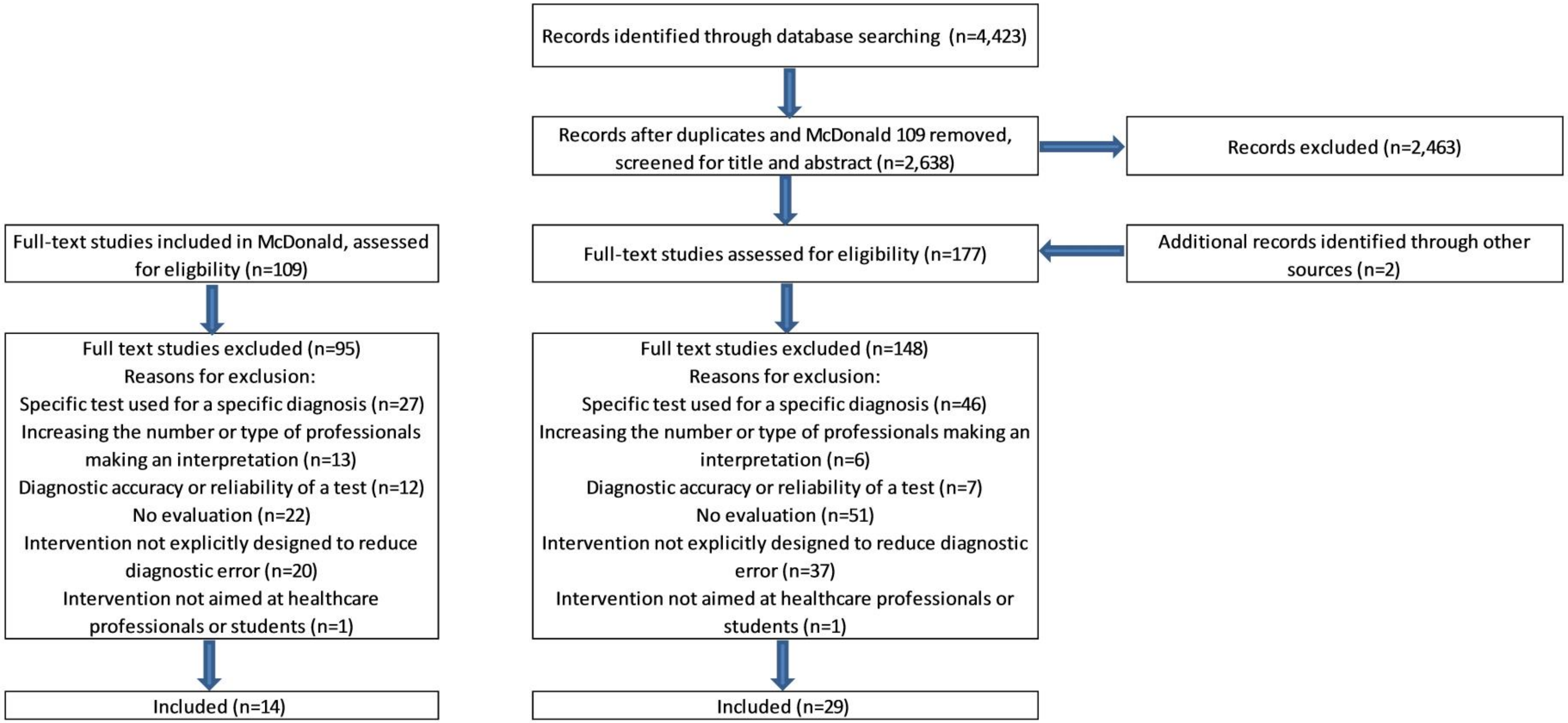

Records identified through database searching (n=4,423)
Records after duplicates and McDonald 109 removed, screened for title and abstract (n=2,638)
Records excluded (n=2,463)
Full-text studies included in McDonald, assessed for eligibility (n=109)
Full-text studies assessed for eligibility (n=177)
Additional records identified through other sources (n=2)
Full text studies excluded (n=95)
Reasons for exclusion:
Specific test used for a specific diagnosis (n=27)
Increasing the number or type of professionals making an interpretation (n=13)
Diagnostic accuracy or reliability of a test (n=12)
No evaluation (n=22)
Intervention not explicitly designed to reduce diagnostic error (n=20)
Intervention not aimed at healthcare professionals or students (n=1)
Full text studies excluded (n=148)
Reasons for exclusion:
Specific test used for a specific diagnosis (n=46)
Increasing the number or type of professionals making an interpretation (n=6)
Diagnostic accuracy or reliability of a test (n=7)
No evaluation (n=51)
Intervention not explicitly designed to reduce diagnostic error (n=37)
Intervention not aimed at healthcare professionals or students (n=1)
Included (n=14)
Included (n=29)

Figure 2

Author (effect)	Rezvvy (large effect)	Sherbino (no effect)	
Error	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising		
Cause of error	Sub-optimal cognitive reasoning		
Type of intervention	Addressed with: education & feedback intervention		
Specific intervention	Specific training programme in diagnostic coding	Cognitive forcing strategy training	
Why the intervention should work (if included in paper)	Diagnostic codes are useful for clinical diagnosis	Analytic (System 2) thinking increases number of diagnoses concidered	
Pre-intervention (intervention development and other requirements before the intervention can be implemented on the ground)	Appropriate design of training programme (curriculum and pedagogy) Staff would be given time to attend training	Appropriate design of training programme (curriculum and pedagogy) Staff would be given time to attend training	Cases may not have been complex enough to benefit from analytic reasoning
INTERVENTION IMPLEMENTATION	**Training programme on ICD-10 coding provided**	**Cognitive forcing strategy training provided**	
Post implementation (before immediate outcome can occur)	Clinician attends training Clinician engages with training Clinician learns from training (ICD-10 coding system for psychiatric diagnoses) Clinician retains learning	Clinician attends training Clinician engages with training Clinician learns from training (how to use analytic (System 2) thinking during diagnostic reasoning process) Clinician retains learning	Poor application of cognitive forcing strategies (lack of learning)
IMMEDIATE OUTCOME	**Clinician uses ICD coding as part of diagnostic process**	**Clinican uses analytic (System 2) thinking during diagnostic reasoning in future clinical practice**	Lack of transfer of cognitive forcing strategies to new cases; clinician overwhelmed by additional circumspection required
Post-immediate outcome (before consequences reach the patient/reduction in diagnostic errors can occur)	Use of ICD coding helps clinician to make correct diagnosis	Full range of possible diagnoses considered Clinician makes correct diagnosis	Experts may be more prone to bias than novices (possible diagnoses remain limited)

Figure 3

Appendix 1: Summary of included studies

Study	Source	Participants	Setting	Country	Design	Intervention description	Intervention category	Error	Cause of error	Definition of outcome/error used to calculate effect	Baseline or Control group outcome (grey=error; white=accuracy)	Post or Intervention group outcome (grey=error; white=accuracy)	Effect size (p-value)	Effect size group
Biffl	109	Physicians	Trauma ICU	USA	Pre-post	Tertiary trauma survey	Process	Sub-optimal weighing during physical examination	System-related	Percentage of patients with a missed injury	2.40%	1.50%	Chi-squared=6.71, p=0.001, Cohen's d=0.254	Small
Chern	109	Physicians	Hospital (Emergency Department)	Taiwan	Pre-post	Education about atypical presentations	Education and feedback	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Lack of knowledge/skill/experience	Percentage of patients with a clinically significant adverse event	0.94%	0.43%	Chi-squared=7.17, p=0.007; Cohen's d=0.438 (Combined)	Small
						Feedback about telephone follow-up of high risk patients	Education and feedback							
Coderre	Repeat	Medical students	Simulation	Canada	Pre-post (type of data randomised)	Provision of additional data and querying of initial diagnosis	Process	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal cognitive reasoning	Percentage of participants with correct diagnosis (combined across types of data provided)	45.4%	82.3%	Chi-squared=79.4, p<0.001, Cohen's d=0.953	Large
Dudley	109	Junior doctors	Hospital	UK	Controlled (not randomised)	Technician report written at time of investigation	Education and feedback	Erroneous clinician interpretation of test	Lack of knowledge/skill/experience	Percentage of reports containing an error (minor disagreement, disagreement or significant disagreement), A&E and medical SHOs combined	52.1%	42.3%	Chi-squared=2.74, p=0.098; Cohen's d=0.220	None
Espinosa	109	Emergency Physicians	Hospital (Emergency Department)	USA	Longitudinal	Review of clinically significant errors in blame-free environment	Education and feedback	Erroneous clinician interpretation of test	Lack of knowledge/skill/experience	Percentages of radiograph interpretations with a false negative finding	3.00%	1.20%	Chi-squared=174, p<0.001, Cohen's d=0.515	Medium
		Radiologists and Emergency Physicians				System re-design	Process		System-related		1.20%	0.30%	Chi-squared=150, p<0.001, Cohen's d=0.771	Medium
Goodacre	Repeat	Senior house officers (Junior doctors)	Simulation of a Hospital Emergency Department	UK	RCT (reports randomised not participants)	Computer-interpretation of investigation results provided to clinicians	Technology	Erroneous clinician interpretation of test	Lack of knowledge/skill/experience	Percentage of ECG interpretations with an error (major or minor)	63.6%	58.4%	Chi-squared=1.42, p=0.233, Cohen's d=0.121	None
Hlabangana	Update	Radiographers	Hospital (Paediatric Department)	South Africa	Pre-post	Poster with most common errors	Education and feedback	Technical errors/poor processing of specimen/test	Lack of knowledge/skill/experience	Mean number of errors per chest radiograph film (post = 1 month after intervention)	4.20	3.23	t=3.634, p<0.001, Cohen's d=0.466 (Combined)	Small
						Crash course about most common errors								
Hopkins	Update	Nurses	Hospital	USA	Pre-post	Training on blood sample collection	Education and feedback	Technical errors/poor processing of specimen/test	Lack of knowledge/skill/experience	Percentage of blood cultures that were contamined (post = quarter following intervention)	3.11%	2.02%	Chi-squared=7.75, p=0.005, Cohen's d=0.245	Small
Hosoe	Repeat	Trainee Endoscopists	Simulation	Japan	Cross-over (random allocation of recordings)	Capsule Endoscopy software providing different methods of viewing recordings	Technology	Erroneous clinician interpretation of test	Lack of knowledge/skill/experience	Median number of missed lesions (false negatives) in capsule endoscopy interpretation	N/A	N/A	Median number of false negatives = 1 for each viewing method, p>0.01; impossible to determine effect size from data presented	None
					Longitudinal	Feedback on previous performance	Education and feedback				N/A	N/A	Mean number of false negatives with each step (approx.): 1.4, 2.5, 0.7, 1.0, 0.6. Impossible to determine effect size or statistical significance from data presented	Unclear
Itri	109	Residents and Fellows	Hospital	USA	Difference in differences (residents vs. fellows; pre-post)	Focused missed-Case Conferences for residents only (fellows act as non-random controls)	Education and feedback	Erroneous clinician interpretation of test	Lack of knowledge/skill/experience	Percentage of musculoskeletal radiograph interpretations (across 31 common injuries) with a major discrepancy	Residents (Int): 18.0%; Fellows (Ctrl): 17.9%	Residents (Int): 6.0%; Fellows (Ctrl): 20.6%	Difference in Differences estimator -0.112 (SE 0.054), t=-2.08, p=0.038; Cohen's d for post error rates only=0.644	Medium
Keijzers	Other	Physicians	Trauma Hospital	Australia	Pre-post	Tertiary trauma survey	Process	Sub-optimal weighing during physical examination	System-related	Perecentage of injuries detected during hospital stay that were missed on initial examination (denominator is total patients, not total missed injuries)	3.80%	4.80%	Chi-squared=0.253, p=0.613; Cohen's d=0.126	None
Kostopoulou Greece	Update	GPs	Simulation of Primary Care	Greece	RCT	Computer diagnostic support system before testing	Technology	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal cognitive reasoning	Mean percentage of correct diagnoses across participants	60%	71%	t=3.19, p=0.002; Cohen's d=0.639	Medium
						Computer diagnostic support system after testing					60%	69%	t=2.75, p=0.007; Cohen's d=0.548	Medium
Kostopoulou UK	Other	GPs	Simulation of Primary Care	UK	RCT	Computer diagnostic support system before testing	Technology	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal cognitive reasoning	Mean percentage of correct diagnoses across participants	63%	69%	t=2.37, p=0.019; Cohen's d=0.337	Small
						Computer diagnostic support system after testing					63%	65%	t=0.74, p=0.462; Cohen's d=0.105	None
Kundel	109	Radiologists	Simulation	USA	Difference in differences (with vs. without feedback using cross-over; pre-post)	Computer pattern recognition	Technology	Erroneous clinician interpretation of test	Sub-optimal cognitive reasoning	Increase in accuracy from initial to second view (area under AFROC curve)	-0.04	0.16	Paired t=40.34, p<0.001; Cohen's d=2.270	Large
Lewis	109	GPs	Primary Care	UK	RCT (patients randomised)	Patient completed questionnaire (PROQSY)	Process	Failure/delay in eliciting critical piece of history data	System-related	Clinical outcomes of patients with possible mental disorder (mean General Household Questionnaire scores/36 at 6 weeks; lower scores are better)	26.6	25.7	t=1.43, p=0.155; Cohen's d=0.160	None
Li	Update	Various	Hospital	China	Pre-post	Computer aided double-signing for samples	Process	Sample mix-up/mislabelled	Sub-optimal attention	Percentage of disqualified samples (post = 1-3 months after intervention)	1.36%	1.19%	Chi-squared=23.8, p<0.001; Cohen's d=0.075 (Combined)	Very small
						Training on sample management and standardized blood sample collection	Education and feedback	Technical errors/poor processing of specimen/test	Lack of knowledge/skill/experience					
						Reference materials on sample collection produced	Education and feedback							
						Improved storage facilities	Process	Technical errors/poor processing of specimen/test	System-related					
						More delivery staff	Process							
Lillo	Repeat	Nurses	Hospital	Spain	Longitudinal	Computer test support	Technology	Failure/delay in performing ordered tests	Sub-optimal attention	Percentage of samples with an error across hematology, coagulation, chemistry and urine samples	0.84%	0.70%	Chi-squared=7.12, p=0.008; Cohen's d=0.097	Very small
						Leaflet explaining blood drawing procedure and explanation of procedure by senior nurse	Education and feedback	Technical errors/poor processing of specimen/test	Lack of knowledge/skill/experience		0.70%	0.38%	Chi-squared=57.5, p<0.001; Cohen's d=0.336	Medium
Mamede	Repeat	Residents	Simulation of Internal Medicine	Netherlands	Pre-post	Structured reanalysis of case findings	Process	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal cognitive reasoning	Mean percentage diagnostic accuracy score on four cases subject to availability bias (previous experience of a similar case; Phase 2 to Phase 3 in the study), across participants, combined first and second years	44.8%	54.3%	t=-1.60, p=0.114; Cohen's d=0.377 (data to enable paired t-test to be undertaken not presented)	None
Monteiro	Update	Residents	Simulation of Medicine Department	Canada	Pre-post	Self-directed reflection	Process	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal cognitive reasoning	Mean percentage diagnostic accuracy score across participants	60.0%	61.0%	t=2.15, p=0.03; Cohen's d cannot be determined (data to verify t-test cannot be determined)	Unclear but possibly very small
Mueller	109	GPs	Primary Care	Germany	Post only with GP confirmation	Patient completed questionnaire	Process	Failure/delay in eliciting critical piece of history data	System-related	Number of health problems uncovered using questionnaire that were previously unknown by the GP	0	Median: 2 (IQR 1-4)	Cannot be determined from the data presented	Unclear
Murphy	Update	Primary Care Providers	Primary Care	USA	RCTs (PCPs randomised)	Reminders	Process	Failure/delay in ordering needed referral	Sub-optimal attention	Percentage of patients with abnormal findings followed-up for diagnostic evaluation by final review (7 months)	52.5%	73.4%	Chi-squared=35.4, p<0.001; Cohen's d=0.511	Medium
Myung	Update	Medical students	Simulation	South Korea	RCT	Enhanced analytical reasoning using structured template	Process	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal attention	Mean percentage diagnostic accuracy score across participants	76.3%	85.0%	t=2.46, p=0.015; Cohen's d=0.355	Small
Nicholl	Repeat	Doctors	Neurology out-patients	UK	Pre-post	Patient feedback	Education and feedback	Failure/delay in eliciting critical piece of history data	Sub-optimal attention	Percentage of missed examinations across all patients in both trusts (3 examinations per patient expected)	31.0%	25.2%	Chi-squared=1.072, p=0.301; Cohen's d=0.156	None
Nishikawa	Repeat	Radiologists	Simulation	USA	Pre-post	Computer test support	Technology	Erroneous clinician interpretation of test	Lack of knowledge/skill/experience	Mean percentage of true positive lesions detected on mammograms across readers	54.9%	60.3%	Paired t=3.91, p=0.006; Cohen's d=1.382	Large
Raab	109	Various/Not stated	Laboratories	USA	Longitudinal	Participation in cross-institution benchmarking programme	Quality improvement	Technical errors/poor processing of specimen/test	System-related	Mean reduction in discordant diagnosis rate for each number of years of participation in programme	N/A	N/A	Mean reductions: 1 year 0.84%, 2 years 0.93%, 3 years 0.97%, 4/5 years 0.99%, p=0.04; Cannot determine effect size from data presented	Unclear but possibly small
Ramirez	Update	Nurses	Intensive Care Unit	Spain	Controlled (not randomised)	Training on blood sample collection	Education and feedback	Technical errors/poor processing of specimen/test	Lack of knowledge/skill/experience	Percentage of blood cultures that were contamined	23%	13%	Chi-squared=10.9, p=0.001; Cohen's d=0.381	Small
Ramnarayan - Paediatrics	109	Junior doctors	4 hospitals (Paediatric Department)	UK	Pre-post	Diagnostic reminder system	Technology	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal cognitive reasoning	Percentage of "unsafe" diagnostic workups (only of cases where system consulted)	45.2%	32.7%	McNemar Chi-squared=13.0, p<0.001; Not possible to calculate Cohen's d	Unclear but possibly small to medium
Ramnarayan - Simulation	Repeat	Various	Simulation	UK	Pre-post					Mean number of diagnostic errors of omission in 12 cases across participants	5.5	5.0	Repeated measures ANOVA p<0.001 (data to calculate F statistic not presented); Cohen's d=0.335	Small

Appendix 1: Summary of included studies

Study	Source	Participants	Setting	Country	Design	Intervention description	Intervention category	Error	Cause of error	Definition of outcome/error used to calculate effect	Baseline or Control group outcome (grey=error; white=accuracy)	Post or Intervention group outcome (grey=error; white=accuracy)	Effect size (p-value)	Effect size group
Rezvyy	Repeat	Psychiatrists	Simulation	Russia	Difference in differences (control vs. intervention; pre-post)	Specific training programme in diagnostic coding	Education and feedback	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal cognitive reasoning	Mean number of correct diagnoses for all cases across participants	Pre: 45.6% Post: 72.4%	Pre: 42.1% Post: 51.3%	p<0.001 for gain in intervention group and comparing post-test scores between groups (data to calculate test statistic not presented); Cohen's d (post-test scores)=1.196	Large
Rondonotti	Update	Capsule Endoscopy readers	Multiple hospitals	Italy	Pre-post	Training including hands-on training and expert tutorial with group feedback	Education and feedback	Erroneous clinician interpretation of test	Lack of knowledge/skill/experience	Percentage of findings detected	35.1%	37.3%	Paired t=0.57, p=0.575; Cohen's d=0.194	None
Rosskopf	Update	Musculoskeletal radiologists	Hospital (Radiology Deparment)	Switzerland	RCT but comparisons pre-post	Quiet working environment	Process	Failed/delayed transmission of result to clinician	System-related	Percentage of reports with any level of discrepancy in diagnostic content	20.8%	8.8%	Chi-Squared=12.5, p<0.001; Cohen's d=0.550	Medium
						Structured report template	Process		Sub-optimal attention		20.8%	20.0%	Chi-Squared=0.05, p=0.824; Cohen's d=0.027	None
Schriger	109	Physicians	Hospital (Emergency Department)	USA	RCT	Patient completed questionniare	Process	Failure/delay in eliciting critical piece of history data	System-related	Percentage of patients who received a psychiatric diagnosis, consultation or referral (assumes that all should do so)	5.10%	7.61%	Chi-squared=0.50, p=0.478; Cohen's d=0.235	None
Segal	Update	Neurologists	Simulation	USA	Pre-post	Computer-based decision support tool	Technology	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Lack of knowledge/skill/experience	Percentage of cases with a diagnostic error	36%	15%	Chi-squared=48.6, p<0.001; Chi-squared=0.638	Medium
Sherbino Trial	Update	Medical students	Simulation	Canada	RCT	Cognitive forcing strategy training	Education and feedback	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal cognitive reasoning	Percentage of participants correctly identifying the second diagnosis on a "search satisficing bias" case	23.9%	31.0%	Chi-squared=0.86, p=0.355; Cohen's d=0.198	None
Sibbald1 - Cardiac	Update	Residents (Junior doctors)	Simulation	Canada	RCT but comparisons pre-post	Checklist after collecting information without return to patient	Process	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Sub-optimal cognitive reasoning	Percentage of doctors with correct diagnosis of cardiac case	44.8%	44.8%	McNemar Chi-squared=0, p=1; Cohen's d=0	None
						Checklist after collecting information with return to patient	Process				47.4%	56.8%	McNemar Chi-squared=7.4, p=0.007; Cohen's d=1.272	Large
Sibbald2 - Experience	Update	Various clinicians	Simulation	Canada	Pre-post	Checklist to correct mistakes in initial diagnosis	Process	Erroneous clinician interpretation of test	Sub-optimal cognitive reasoning	Mean total number of errors (omitted and incorrect diagnoses) in all ECG cases across participants	26.5	24.9	Repeated measures ANOVA F=12.2, p=0.001; Cohen's d=0.201	Small
Sibbald3 - Experts	Update	Physicians (experts)	Simulation	Canada	Pre-post	Verification stage (no checklist)	Process	Erroneous clinician interpretation of test	Sub-optimal cognitive reasoning	Mean number of errors per ECG (omitted and incorrect diagnoses) across participants	1.66	1.63	t=0.13, p=0.896 (data to calculate paired t-test statistic not presented); Cohen's d=0.020	None
						Checklist to correct mistakes in initial diagnosis					1.51	1.21	t=1.41, p=0.160 (data to calculate paired t-test statistic not presented); Cohen's d=0.211	None
Tsai	Repeat	Residents	Simulation	USA	RCT (cross-over)	Computer-interpretation of investigation results provided to clinicians	Technology	Erroneous clinician interpretation of test	Lack of knowledge/skill/experience	Mean percentage of findings correctly interpreted across participants (regardless of accuracy of computer system)	48.9%	55.4%	Paired t cannot be determined from data presented, p<0.001; Cohen's d=0.628	Medium
Tudor	Repeat	Physicians	Simulation of Radiology Department	UK	Pre-post	Individual feedback on image interpretation	Education and feedback	Erroneous clinician interpretation of test	Lack of knowledge/skill/experience	Percentage accuracy of reporting across radiologists	82.2%	88.0%	Paired t=2.54, p=0.032; Cohen's d=0.803 Results for each radiologist had to be read from a graph	Large
Turner	Update	GPs	Primary Care	UK	Pre-post	Computer test support	Technology	Sample mix-up/mislabelled	Sub-optimal attention	Percentage of samples with any error	1.25%	0.21%	Chi-squared=1644, p<0.001; Cohen's d=0.981	Large
Weatherburn	109	Senior house officers (Junior doctors)	Hospital (Emergency Department)	UK	Pre-post	Computerised version of images	Technology	Erroneous clinician interpretation of test	System-related	Percentage of radiographed patients with any level of misdiagnosis	1.51%	0.65%	Chi-squared=13.7, p<0.001; Cohen's d=0.464	Small
Wellwood	109	Senior house officers (Junior doctors)	Hospital (Emergency Department)	UK	RCT	Standardised data collection forms	Process	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Lack of knowledge/skill/experience	Percentage of initial diagnoses that were incorrect	41%	35%	Unable to determine from data presented (percentages are approximate as read from a graph)	Unclear
					RCT of incremental effect (data for pre-post only)	+Computer-based decision support tool	Technology				35%	32%		
					Pre-post	+Monthly feedback	Education and feedback				32%	29%		
Wexler	109	Physicians	Hospital (Paediatric Department)	USA	Non-randomised controls (odd/even day admissions)	Computer-based decision support tool	Technology	Failure/delay in considering the correct diagnosis; sub-optimal weighing/prioritising	Lack of knowledge/skill/experience	Mean time to diagnosis (days)	2.8	1.9	p>0.05; test statistic and Cohen's d cannot be calculated from data presented	None

Appendix 2: Logic Models

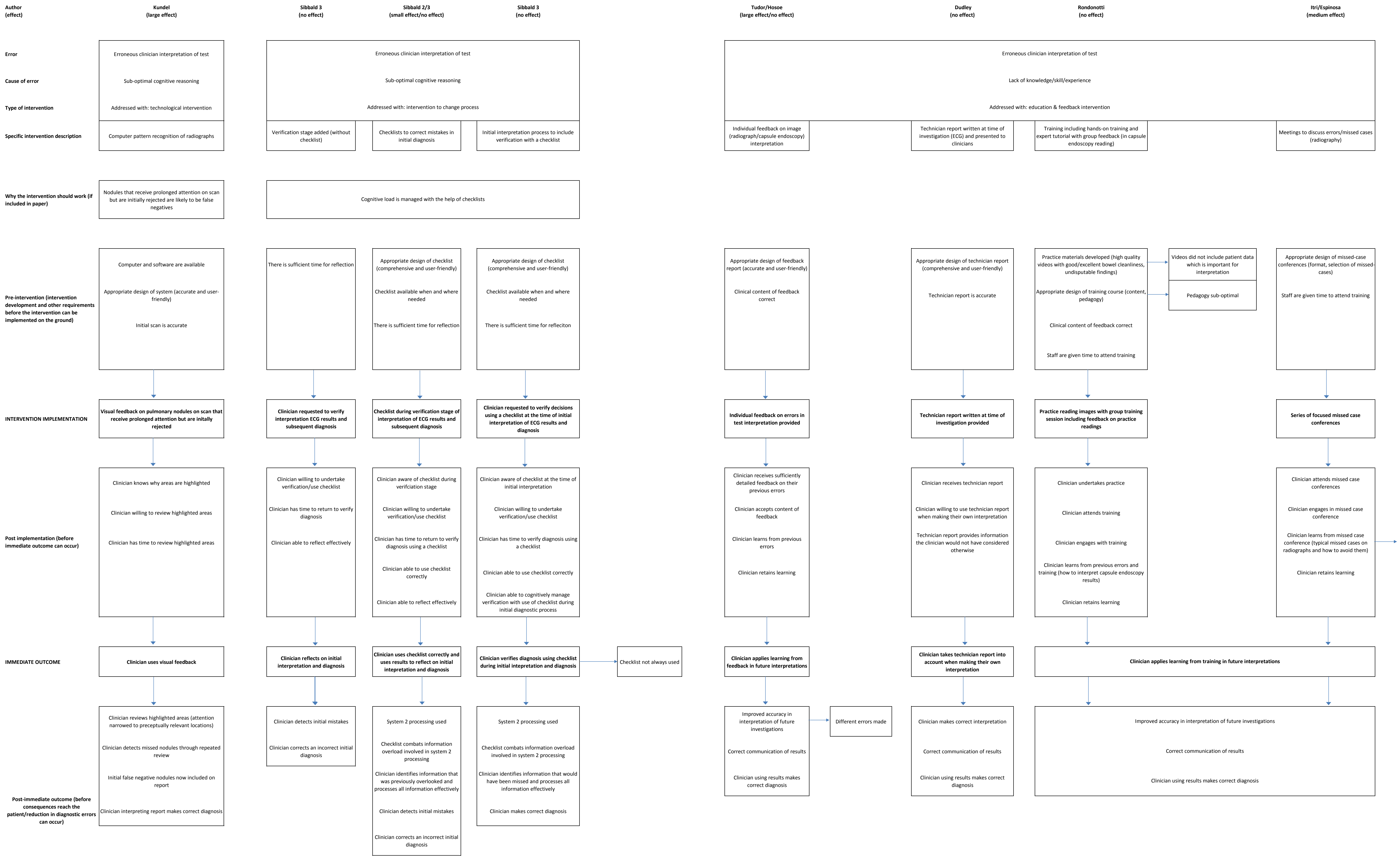
Author (effect)	Lillo (medium effect)	Turner (large effect)		Li (very small effect as multifaceted interventon)
Error	Failure/delay in performing ordered tests	Sample mix-up/mislabelled		Sample mix-up/mislabelled
Cause of error	Sub-optimal attention	Sub-optimal attention		Sub-optimal attention
Type of intervention	Addressed with: technological intervention	Addressed with: technological intervention		Addressed with: intervention to change process
Specific intervention description	Computer test support	Computer test support		Computer-aided double-signing for samples
Why the intervention should work (if included in paper)				
Pre-intervention (intervention development and other requirements before the intervention can be implemented on the ground)	Computer and printer are available Appropriate design of computer system (of labels and instructions)	Computer, printer and internet access are available Appropriate design of computer system (of labels and instructions)		Computer and software are available Appropriate design of computer system (all required information included in reminder)
INTERVENTION IMPLEMENTATION	**Computer-printed custom labels and instructions to correlate labels to test tube**	**Electronic requesting of tests (labels printed and detail of tube(s) required provided)**		**Computer-facilitated double-signing system**
Post implementation (before immediate outcome can occur)	Clinician aware of computer system Clinician willing to use computer system Clinician able to use computer system correctly	Clinician aware of electronic requesting Clinician willing to use electronic requesting Clinician able to correctly request tests electronically	New staff not informed/training gains decline over time	Clinician aware of system Clinician willing to use system Clinician able to use system correctly
IMMEDIATE OUTCOME	**Clinician uses computer system**	**Clinician uses electronic system to order tests**		**Clinicians uses computer-facilitated double-signing system**
Post-immediate outcome (before consequences reach the patient/reduction in diagnostic errors can occur)	Technology works as intended Clinician able to draw all samples required (depends on patient condition) Clinician able to add correct label to each sample Samples arrive at lab in good condition Samples are tested correctly Correct communication of test results to clinician Clinician interpeting results makes correct diagnosis	Technology works as intended Clinician able to draw all samples required (depends on patient condition) Clinician able to add correct label to each sample Samples arrive at lab in good condition Samples are tested correctly Correct communication of test results to clinician Clinician interpeting results makes the correct diagnosis	Printer did not produce labels correctly	Technology works as intended Staff communicate all required information related to the samples Samples arrive at lab in good condition Samples are tested correctly Correct communication of results to clinician Clinician interpreting results makes correct diagnosis

Author (effect)	Hlabangana (small effect)		Ramirez/Hopkins (small effect)	Li (very small effect as multifaceted interventon)		Lillo (very small effect)
Error	Technical errors/poor processing of specimen/test					
Cause of error	Lack of knowledge/skills/experience					
Type of intervention	Addressed with: education & feedback intervention					
Specific intervention description	Poster with most common errors (on radiographs)	Crash course about most common errors (on radiographs)	Training on blood sample collection	Training on sample management and standardized blood sample collection	Reference materials on blood sample collection produced	Leaflet explaining blood drawing procedure and explanation of procedure by senior nurse
Why the intervention should work (if included in paper)						
Pre-intervention (intervention development and other requirements before the intervention can be implemented on the ground)	Appropriate design of poster (clear content, attrative lay-out)	Appropriate design of training (content, pedagogy) Staff would be given time to attend training			Appropriate design of materials (clear content, attractive lay-out)	
INTERVENTION IMPLEMENTATION	**Poster with images of most common radiographic technical errors**	**Crash course on most common radiographic technical errors**	**Training on correct collection of blood culture specimens for phlembotomists**	**Training on sample collection and management**	**Materials on sample collection**	**Leaflet with correct sample procedure**
Post implementation (before immediate outcome can occur)	Clinician aware of poster Clinician has time to read poster Clinician processes information on common radiographic errors from poster correctly Questions are asked when information is not understood Poster not taken down	Clinican aware of training Clinican attends training Clinician engaged in training Clinician learns from training (how to avoid common errors/how to take and manage samples correctly) Learning is retained			Clinican aware of materials/leaflet Clinical has time to read materials/leaflet Information on sample collection from reference materials processed Questions are asked when information is not understood Materials/leaflet always accessible	
IMMEDIATE OUTCOME	**Clinician aware of most common errors**		**Clinician aware of correct sample procedure**			
Post-immediate outcome (before consequences reach the patient/reduction in diagnostic errors can occur)	Radiographs taken correctly Radiograph interpreted correctly Correct communication of results to clinician Clinicianmakes correct diagnosis		Samples drawn and labelled correctly Samples arrive at lab in good condition Samples are tested correctly Correct communication of results to clinician Clinician interpreting results makes correct diagnosis			

Appendix 2: Logic Models

Author (effect)	**Li (very small effect as multifaceted intervention)**		**Raab (unclear effect)**	**Nicholl (no effect)**	**Weatherburn (small effect)**	**Rosskopf (medium effect)**	**Rosskopf (no effect)**
Error	Technical errors/poor processing of specimen/test		Technical errors/poor processing of specimen/test	Failure/delay in eliciting critical piece of history data	Erroneous clinician interpretation of test → Insufficient headroom for improvement (low baseline error rate)	Failed/delayed transmission of result to clinician	Failed/delayed transmission of result to clinician
Cause of error	System-related		System-related	Sub-optimal attention	System-related	Sub-optimal attention	System related
Type of intervention	Addressed with: intervention to change process		Addressed with: quality Improvement intervention	Addressed with: education & feedback intervention	Addressed with: technological intervention	Addressed with: intervention to change process	Addressed with: intervention to change process
Specific intervention description	Improved storage facilities	More delivery staff	Participation in cross-institution benchmarking programme	Patient feedback (on use of instruments)	Computerised version of images	Quiet working environment	Structured report template
Why the intervention should work (if included in paper)				To make an accurate diagnosis all instruments should be used during examination	View of radiograph is improvend when images are manipulated		
Pre-intervention (intervention development and other requirements before the intervention can be implemented on the ground)	Samples are collected correctly Samples are labelled appropriately Storage facilities are in working order	New staff recruited New staff able to handle samples correctly (may require effective training)	Appropriate design of data management system (so that data are useful) All institutions provide timely data accurately and honestly Samples are collected correctly Samples are labelled appropriately	Appropriate design of patient questionnaire Patient questionnaire available when and where needed Patient completes questionnaire and does so accurately → Recall bias	Clinicians forced to use computerised images (hard copy films not available to clinicians to view) Appropriate design of software for manipulation of images	Space for a quiet room is available	Other methods of reporting removed so no alternative process Appropriate design of structured report template (comprehensive and user-friendly)
INTERVENTION IMPLEMENTATION	**New sample storage facilities provided**	**More sample delivery staff**	**Long-term participation in cross-institutional benchmarking programme**	**Feedback about the use of all instruments provided to clinician**	**Soft-copy radiographic images can be manipulated online and discussed by telephone with radiologist**	**Quiet environment provided for report writing**	**Structured report template provided**
Post implementation (before immediate outcome can occur)	Clinician aware that facilities are available Clinician willing to use facilities Clinician able to use facilities correctly	Clinician able to request delivery staff Delivery staff available when required	Data are received Data are read Data are analysed Decision to take action made Appropriate interventions developed	Clinician receives and reads email with sufficiently detailed feedback on patient-reported use of instruments Clinician accepts content of feedback Clinician decides to use all instruments in future consulations All instruments available and in working order in future consultations → Instruments required are not available Clinician able to use all instruments correctly	Clinician aware that images can be manipulated Clinicians willing to manipulate images Clinician has time to manipulate images Clinician able to manipulate images effectively	Clinician aware of quiet room and knows where it is Clinicians willing to use quiet room	Clinician able to use template correctly → Adjustment period required for clinicians to learn new process
IMMEDIATE OUTCOME	**Sample storage facilities used correctly (sample kept in better condition)**	**New staff reduce time taken to get sample to lab**	**Institution implements appropriate improvement interventions (not specified)**	**Clinician accurately uses all instruments in future examinations**	**Clinician manipulates images effectively**	**Clinician uses quiet room**	**Clinician uses structured report template**
Post-immediate outcome (before consequences reach the patient/reduction in diagnostic errors can occur)	Samples arrive at lab in good condition Samples are tested correctly Correct communication of results to clinician Clinician interpreting results makes correct diagnosis		Interventions are effective in reducing sample testing errors Correct communication of test results to clinician Clinician interpeting results makes correct diagnosis	Clinician obtains useful information Clinician able to interpret results to make correct diagnosis	Clinician obtains useful information Clinician able to interpret results to make correct diagnosis	Quiet room is quiet and clinician not interrupted Clinician is focused when writing report Clinician writes report accurately and completely Clinician interpeting report makes the correct diagnosis	Clinician writes report accurately and completely Clinician interpeting report makes the correct diagnosis

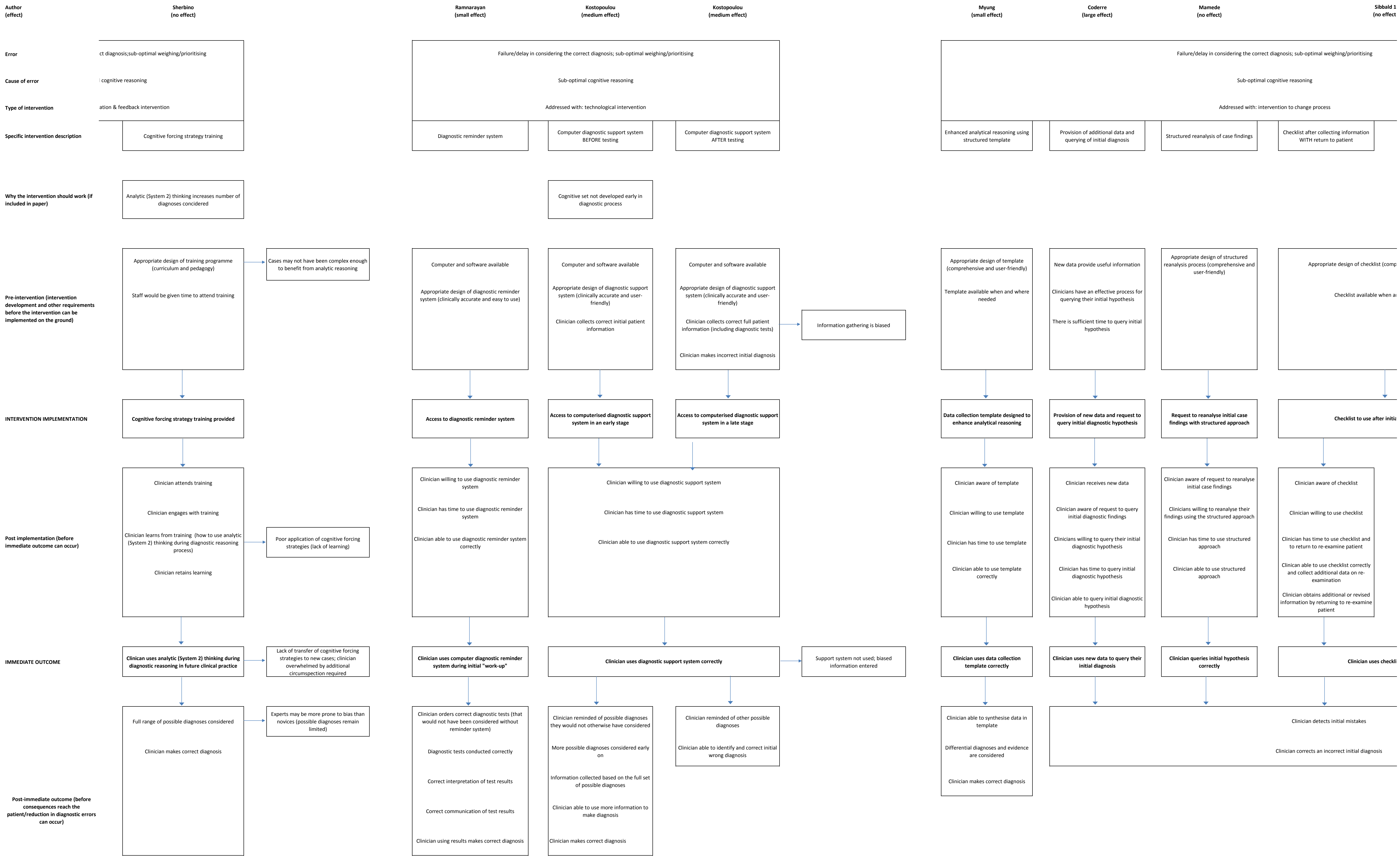
Appendix 2: Logic Models



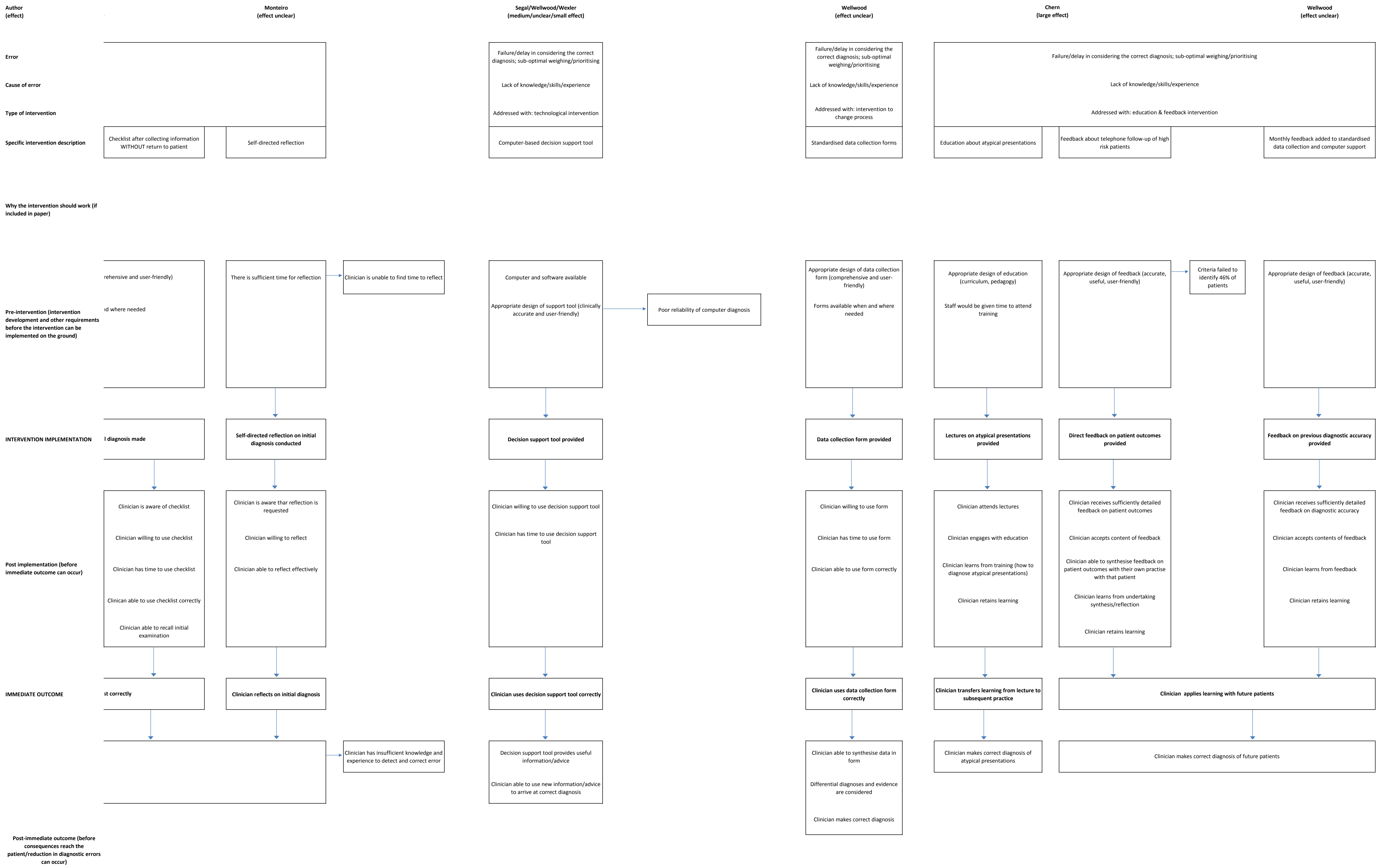
Appendix 2: Logic Models

Author (effect)		Hosoe (effect unclear)		Nishikawa (large effect)	Tsai /Goodacre (medium effect/no effect)		Espinosa (medium effect)	Murphy (medium effect)	Rezvyy (large effect)
Error		Erroneous clinician interpretation of test					Erroneous clinician interpretation of test	Failure/delay in ordering needed referral	Failure/delay in considering the corre
Cause of error		Lack of knowledge/skills/experience					System-related	Sub-optimal attention	Sub-optima
Type of intervention		Addressed with: technological intervention					Addressed with: intervention to change process	Addressed with: intervention to change process	Addressed with: educ
Specific intervention description		Software to help clinicians read images (capsule endoscopy)		Computer-aided detection of mammography screening	Computer-interpretation of investigation (ECG) results provided to clinicians		Structured reporting process (radiography results)	Reminders	Specific training programme in diagnostic coding
Why the intervention should work (if included in paper)								All abnormal results need follow-up	Diagnostic codes are useful for clinical diagnosis
Pre-intervention (intervention development and other requirements before the intervention can be implemented on the ground)		Computer and software available Appropriate software design (accurate identification of positive findings, user-friendly)	Software has a high false negative rate	Computer and software available Appropriate software design (accuracte identifcation of positive findings, user-friendly)	Computer and software available Appropriate design of report giving advice (user-friendly, comprehensive)	Software provides incorrect advice	Appropriate design of process to correctly identify problems	Human and technological resources available to run reminder generation system and communicate results Appropriate design of reminder system (sensitive/specific, user-friendly)	Appropriate design of training programme (curriculum and pedagogy) Staff would be given time to attend training
INTERVENTION IMPLEMENTATION		**Software selects most important images to be viewed by clinician**		**Computer-aided cancer detection report provided**	**Computer generated report concerning interpretation of ECG results**		**Process redesign to designate initial, checking and patient notification responsbilities**	**Reminders provided (repeated if no action)**	**Training programme on ICD-10 coding provided**
Post implementation (before immediate outcome can occur)	New staff unaware of typical errors (mitigated by manatory study of file of missed cases)	Clinician aware of software Clinician willing to use software Clinician able to use software		Computer generated report provided to clinician Clinician willing to use computer generated advice Clinician has time to use report Report provides information clinician would not have considered otherwise		Clinician ignores report	Clinician aware of the new process and its requirements Clinician willing to use new process Clinician has time to use new process	Clinician receives reminder (repeat) Clinician reads reminder (repeat) Clinician decides to take action to recall the patient	Clinician attends training Clinician engages with training Clinician learns from training (ICD-10 coding system for psychiatric diagnoses) Clinician retains learning
IMMEDIATE OUTCOME		**Clinician uses software to review images**		**Test interpreted with the help of computer generated report**			**Clinician uses new process as intended**	**Clinician recalls patient for follow-up (successful contact made)**	**Clinician uses ICD coding as part of diagnostic process**
Post-immediate outcome (before consequences reach the patient/reduction in diagnostic errors can occur)		Clinician directed to images with positive findings Clinician less likely to miss a positive finding Correct interpretation of test results Correct communication of test results Clinician using results makes correct diagnosis		Correct interpretation of test results Correct communication of test results Clinician using results makes correct diagnosis			Reporting clinician able to complete original report Checking clinician able to detect and correct errors Patients are recalled for follow-up where required (successful contact made) Patients recalled attend follow-up Clinician makes correct diagnosis (at follow-up if required)	Patient attends recall Clinician makes correct diagnosis at recall	Use of ICD coding helps clinician to make correct diagnosis

Appendix 2: Logic Models



Appendix 2: Logic Models



Appendix 2: Logic Models

Author (effect)	**Lewis/Mueller/Schriger (no effect/effect unclear/no effect)**		**Biffl/Keijzers (small effect/no effect)**	
Error	Failure/delay in eliciting critical piece of history data		Sub-optimal weighing during physical examination	
Cause of error	System-related		System-related	
Type of intervention	Addressed with: intervention to change process		Addressed with: intervention to change process	
Specific intervention description	Patient-completed questionnaire (about symptoms/problems)		Tertiary trama survey	
Why the intervention should work (if included in paper)				
Pre-intervention (intervention development and other requirements before the intervention can be implemented on the ground)	Appropriate design of questionnaire (valid and user-friendly) Patient completes questionnaire and does so accurately		Appropriate design of survey (includes examinations required to identify diagnoses missed on admission; user-friendly)	
INTERVENTION IMPLEMENTATION	**Results of questionnaire provided to clinician**		**Tertiary trauma survey process implemented**	
Post implementation (before immediate outcome can occur)	Clinician receives and reads results	Clinician did not read/ignored results	Clinician willing to conduct survey	Lack of governance to encourage use; clinicians fear loss of autonomy
	Clinician accepts contents of results		Clinician able to conduct survey	Staff turnover means new staff not aware of requirements
	Clinician obtains useful information		Clinician has time to conduct survey	High workload/external pressures reduce time available
	Clinician decides to take action	Clinician has adverse beliefs related to consequences for patient's insurance/employment and/or lack of ongoing care options	Patient ambulatory and conscious at time of survey	
IMMEDIATE OUTCOME	**Clinician acts on results of questionnaire and recalls patient**		**Clinician conducts survey**	
Post-immediate outcome (before consequences reach the patient/reduction in diagnostic errors can occur)	Clinician has time to see patient again		Clinician obtains useful information	
	Patient attends follow-up	Patients did not attend follow up	Clinician able to interpret results	
	Clinician makes correct diagnosis		Clinician makes correct diagnosis	