GP Benchmark: Engineering a Crowd-Sourcing Platform for Real-Time Understanding of Personality and Cognitive Biases in Clinical Error

Wesley Hutchinson
Lancaster Medical School
Lancaster University
Lancaster, UK
w.hutchinson@lancaster.ac.uk

Sumi Helal
Computing and Communications
Lancaster University
Lancaster, UK
s.helal@lancaster.ac.uk

Christopher Bull
Computing and Communications
Lancaster University
Lancaster, UK
c.bull@lancaster.ac.uk

***Abstract*—Errors in medicine are a significant problem, highlighted as a global safety priority. General Practice is one clinical arena where error is more likely due to clinical decisions being made on a background of clinical complexity, undifferentiated symptoms and diseases, and multiple other factors as yet unquantified. Interventions designed to reduce error are either underutilised, untested, fail to produce lasting results, are designed on inadequate knowledge, or have failed to appreciate the interaction of multiple factors, both cognitive and systemic. We present a potential solution, in the form of GP Benchmark. GP Benchmark is an online simulation environment and tool designed to test clinical decision making in a group of practicing General Practitioners. Its aim is to address two pressing requirements: 1) the need to capture clinical decision making in real-time, in the context of personality, cognitive bias and environmental factors, and 2) the need to provide a validated platform that models the clinical environment so future intervention decisions may be tested without risking patient safety. We highlight the requirements satisfied for implementing GP Benchmark, the plans for validation, and discuss how GP Benchmark will be used to identify further requirements necessary to develop the environment into a tool for testing clinical decision support systems and error prevention strategies.**

***Index Terms*—Medicine, Cognition, Cognitive Informatics, Systems Simulation, Requirements Engineering**

I. Introduction

Error has been highlighted as a global safety priority and a significant problem in medicine [1], posing a significant threat to the health and wellbeing of both the patient and the practitioner [2]. With up to an estimated 6 million clinically significant errors occurring each year in the UK [3], General Practice is a speciality particularly susceptible to error. It encapsulates the core elements of a complex system – uncertainty and unpredictability [4]; patients often present early in the disease course with poorly differentiated symptoms [5]–[7].

The wrong diagnosis occurs 10-15% of the time [8]. Decision making lies at the heart of General Practice and there are multiple points throughout a clinical consultation which demand a decision from the clinician and the possibility of an error being made [9]. Diagnostic error can be broadly divided into deficiencies of knowledge, data gathering, information processing and verification [8], [10]. Information processing and verification present the largest areas of error, with deficiencies in knowledge representing a very small proportion [8], [11]–[13].

Cognitive bias has gained focus as an area of research with potential to explain some of the errors observed in clinical practice. However, we would argue that investigation of cognitive bias in the clinical arena has been insufficient, relying on researchers to investigate errors retrospectively through reflective practice and assign cognitive biases based on interpretation of documented events.

Graber et al. [8] reviewed the medical records of 100 patients and found that 74% of errors were due to some failure of cognitive processing and 65% were system-related. They highlighted that most errors in medicine result from some combination of both cognitive and system-related factors. However, stating there are multiple factors contributing to the resulting error is one thing, observing these multiple factors, in the precise manner in which they contribute to the error, is significantly more difficult. Identifying these factors is difficult enough after the event has occurred, it is unquestionably more challenging, but arguably more important, to identify them in the case of error prediction and mitigation.

We present our vision as a step toward addressing this problem. We have created GP Benchmark as a means of developing dynamic and integrated, cognitive and environmental models; more definitively, GP-Patient-decision models, to explore and understand the possible sources of error in healthcare, specifically General Practice. Modelling requires integration of a clinical-consultation-process model, a GP cognitive process architecture, a patient model with bias-triggering elements, and models of both environmental and systemic constructs experienced within a consultation.

This modelling effort was motivated by initial exploration activities and dialogue with practicing GPs. When error occurs in General Practice, a root-cause-analysis, commonly known as a significant event analysis (SEA) is undertaken and presented to peers to identify learning opportunities. There were two common elements from the discussion about SEA:

1) Similar errors appeared in an almost cyclical nature

every few years, highlighting that learning gained from the process was finite.

2) Learning was often at the mercy of hypothetical analysis, i.e., if event A had happened in situation B, then the outcome may have been different. For those undertaking the SEA, significant doubts remained regarding whether the error could truly be mitigated in the future.

This paper describes the inherent problems of previous research exploring the area of error in medical practice. In section 2, we highlight current understanding of factors influencing the incidence of clinical error. Section 3 outlines the requirements identified in the development of GP-Benchmark; an online platform created to gain further understanding of the complexity of the Doctor-Patient clinical encounter. Section 4 outlines the necessary requirements, to be obtained via GP-Benchmark, to model the interaction of factors contributing to error. In section 5, we outline the vision of GP-Benchmark, its real-time, real-world application, together with some potential challenges.

II. Background

With poor decision-making estimated to account for 75% of medical errors [14], understanding how clinical decisions are made and how they may be influenced represents an important area of research. Cognitive bias is increasingly recognised as an important factor in decision making errors in General Practice [9], [14]. It can be defined as the failure of a heuristic model but may also be defined as a cognitive predisposition without concern for accuracy [14], [15].

Multiple biases have been identified and cited as factors contributing to diagnostic error [14], [16]–[19]. Previous research approaches to mitigate these biases include interventions helping the individual reflect on their diagnostic process, increase clinical knowledge, use second opinions, or have required the user to utilise specific diagnostic software [20]. Despite this body of research, some over 20 years old, their application to medical education and routine clinical practice remains limited [20], [21].

There are some significant flaws in earlier attempts to reduce error that focus on cognitive bias, including:

- failures to address the multi-factorial nature of how errors arise.
- cognitive biases are researched based on the most statistically common, rather than most clinically relevant.
- poor understanding of when and how cognitive bias arise.
- failures to appreciate an individual clinician's variability and susceptibility to different biases.

Whilst research specifically focusing on errors of medical decision-making has been increasing over the last 60 years, there are clear environmental variables, such as tiredness, information overload, clinical knowledge gaps or poor knowledge of the patient, that make errors more likely to occur [22]–[26]. Polypharmacy, frailty, complex and multiple medical conditions, fragmented reporting systems, increased patient demand, variations in training and experience, fatigue, depression, and burnout all contribute to the risk of error and are all factors prevalent in today's clinical arena [26]–[28].

Despite the growing acceptance as a factor in medical error, cognitive bias is a challenging area to research [15]. There is a distinct lack of high-quality, granular data on the prevalence of error and the effect of cognitive bias, reflecting the logistical difficulties of observing what is an internal, subconscious process [15]. To date, approaches in understanding this area have relied heavily on reviewing notes of deceased patients or self-reflection of errors after the event has occurred [8], [14], [18], [20]. The limitation of this approach is a lack of clarity and appreciation of the cognitive process at the point of decision, relying on interpretation of decisions and actions after the event.

Currently, any insights or learning points generated from errors in General Practice are limited to the individual or immediate team. There is no clear mechanism to disperse this information to other clinicians or integrate insights into clinical practice and change behaviour. Given the paucity of evidence and difficulties in recording the cognitive process involved in diagnostic, treatment and management decisions it is imperative to develop solutions which may help clinicians identify and reduce error in their medical practice.

III. Developing Requirements for GP-Benchmark

Developing requirements for GP Benchmark initially involved an extensive review of the literature surrounding error in clinical practice and methods to reduce these. In our overarching aim to mitigate error in clinical practice, it quickly became apparent that current research into this area had modelled the area too simplistically. Interventions to reduce error have focused on individual elements of the system or the clinician that are thought to be the most relevant, rather than addressing the complex interaction of these factors [20], [30]. Whilst some of the suggested interventions show promise, testing has been non- or quasi-experimental [20], [30].

Consultation with practicing GPs highlighted a recurring dichotomy; the need to disseminate learning derived from previous errors versus the limits of GP time and ability to assimilate learning into clinical practice. The current approach is the use of Clinical Decision Support Systems (CDSS), designed to present the user with information at the point of decision. These systems have undergone extensive requirements engineering activities to be used in clinical practice, but real-world improvements have had mixed results [31]. We conducted a survey exploring the current stakeholder's perceptions of their Electronic Health Record (EHR) and CDSS. Whilst feedback was mostly positive, the scope of current CDSS explored is limited to searching for medical knowledge, requiring the user to identify a deficit and then query the CDSS. Extending the performance and scope of these systems in real world settings, highlighted a further requirement, the provision of a validated environment that could test their use and impact in clinical practice.

We realised that understanding the clinical consultation process that is followed by clinicians would prove imperative

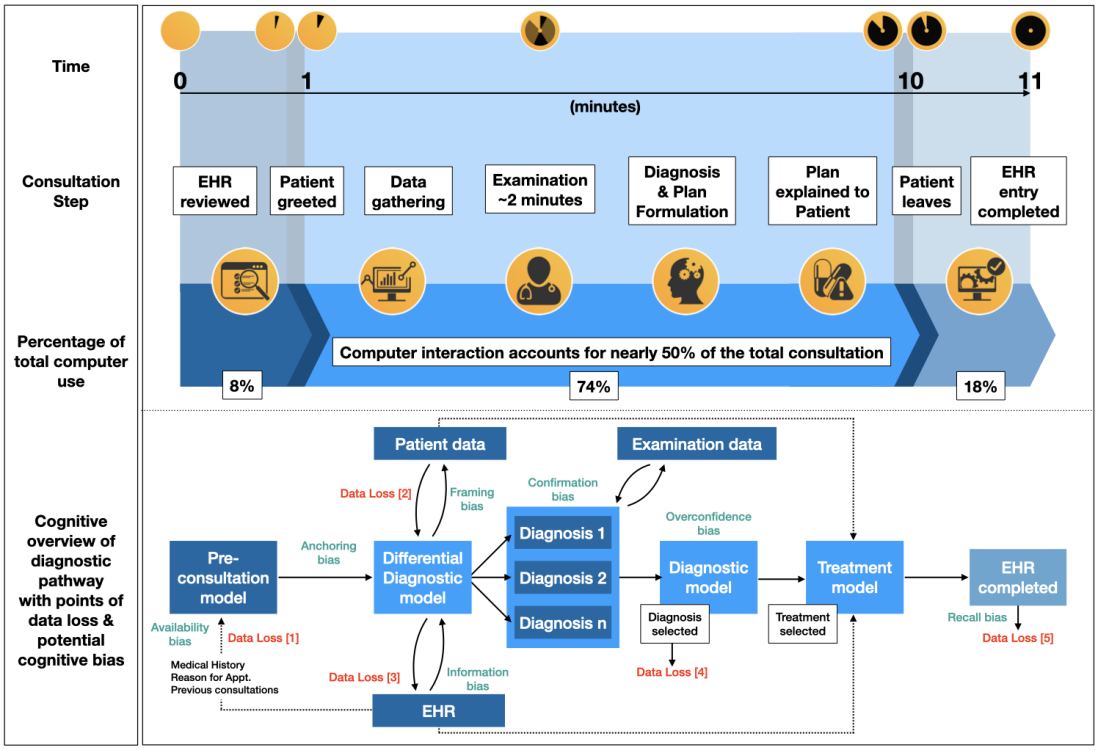

Fig. 1. Infographic representation of a GP consultation model for decision points in a diagnostic pathway, illustrating potential points of bias and data loss within the process, combined with observational data from Kumarapeli & de Lusigan [29], demonstrating time constraints and computer interaction.

to identifying requirements for modelling the clinical environment and measuring decision points and outcomes. In this paper we provide an outline of our current consultation model and the initially identified points of potential cognitive bias and data loss (Fig. 1). This outline provides the foundation to further modelling the environment for testing, capturing and assessing clinical decision making.

The initially identified biases, based on our consultation model, are [32]:

- **Anchoring bias**: a tendency for decisions to be made on previously presented data (the "anchor")
- **Availability bias**: a susceptibility to reach judgements based on data, decisions or outcomes more readily available in recent memory.
- **Confirmation bias**: a predisposition to find, notice and remember information which fits with pre-existing ideas.
- **Framing effect**: based on how information is presented, different conclusions may be presented.
- **Information bias**: the variability in how key information is measured, collected, or interpreted.
- **Recall bias**: a subset of information bias, concerned with the accuracy in which information is remembered
- **Overconfidence bias**: overestimation of one's own ability to successfully make a correct decision.

Fig. 1 is split into upper and lower parts with time referenced from left to right, the average GP consultation in the UK is under 11 minutes [29]. Upper elements outline, in broad terms, actions undertaken in the consultation, from reviewing the EHR prior to seeing the patient, to gathering information, examining and making a diagnosis, before formulating a management plan and completing the EHR entry. Observations from Kumarapeli & de Lusigan [29] identified the timings of these events and that approximately 50% of this time is spent interacting with the computer.

Lower components of Fig. 1 demonstrate potential areas of cognitive bias, data loss and error. Specified biases are based on most likely hypothesised biases, but in most cases there is likely to be multiple factors influencing decisions.

The clinical consultation starts with the generation of a 'pre-consultation' model developed from information in the EHR and the identified reason for patient attendance. Data is lost at this point due to multiple factors: how the attendance reason fits with current known medical problems, the most recent entries in the EHR and the clinician's clinical knowledge and most accessible experiences. The initial model is used to generate the differential diagnosis model. The clinician pattern matches information presented to them to known disease and illness models to identify a list of possible diagnoses. How the information is accessed and assimilated from the patient and the EHR is subject to subconscious filtering and exclusion, biasing data used to generate the list. An analogy of the diagnostic process could be that of using a sieve. Lots of information and data is presented to the clinician. Data that fits together to form a diagnosis is held within the sieve, but everything else is filtered out. Filtered data is often not recorded and is then lost to future consultations.

One diagnosis is preferentially chosen based on its match to the identified disease model. Other differentials and data may be documented in cases where significant diagnoses must be excluded, with relevant negative findings included. On completion, a record of the visit, as interpreted by the clinician and subject to recall bias, is entered into the EHR and further data is lost. By excluding information deemed irrelevant by the clinician, the EHR will be ready to potentially bias the next user, propagating the cognitive bias to future visits.

IV. The Focus of GP-Benchmark

As a finished environment, GP Benchmark will be an online platform designed to model, simulate and test human cognition and performance within the context of the Doctor-Patient clinical interaction. It will provide a unique architecture to test and evaluate CDSS and interventions to mitigate error. GP Benchmark is required to have a dual role initially, assessing decision making, but also gathering further requirements through user interaction and feedback.

Using simulation, whether through clinical scenarios or simulated patients, has the unmatched utility to expose GPs to a greater number and variety of topics, provide controlled and reproducible scenarios and has the potential to improve clinical reasoning [20]. However, it has not yet been significantly explored as a tool for assessing decision-making related to clinical practice. Whilst it remains unproven that simulation can replace clinical experience as a learning tool, it has been demonstrated that immediate and focused feedback can improve performance in selected settings [20]. Fig. 2 shows the user's dashboard view, providing an area to reflect on simulated clinical cases they have completed.

One of the first decisions we made when deciding to model the clinical consultation was a decision about scope. There are multiple factors involved in the activity of diagnostic deduction and developing a management plan. Following reflection of the process model described above, understanding the clinical environment and discussion with practicing GPs, we made the compromise, initially, to limit the scope of the simulation by modelling the activity of information processing and decision-making. We have chosen to model only the patient data received and accessed in a consultation, removing both time and physical factors of the environment. We thought this decision was a good balance between accuracy and relevance, as much of the factors thought to influence error are based on hypothesis [33].

A second, important aspect of scope was to select the cognitive biases on which we would focus our efforts. Using the designed model in Fig. 1, those biases most recognised in the literature provided the initial focus for incorporation and investigation. In this light, we considered both the strengths and weaknesses of this approach with alternative approaches from prior research. We made the decision to use our model as a foundation that would be open to previously unforeseen factors and biases which may become apparent. We have decided to concentrate on the dynamic decision-making aspects of the task in its closed-loop context. We focus on factors

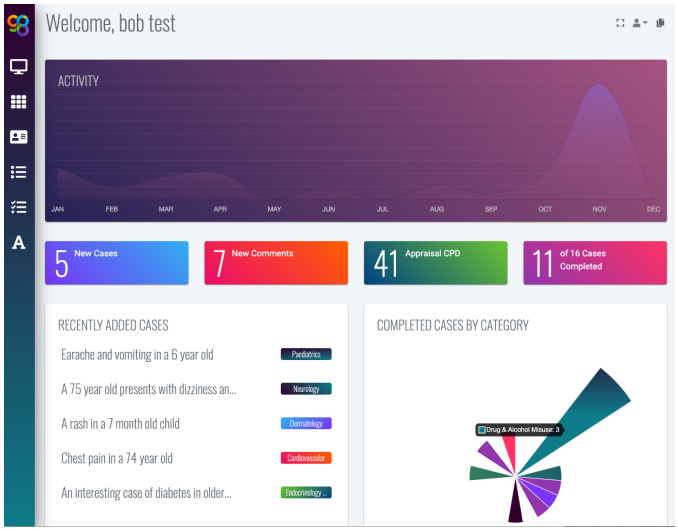

Fig. 2. Dashboard view of online platform.

that may result from the interaction of the GP with imperfect, incomplete and biased information.

Initial modelling of the human in the loop, the GP, is simplistic as it represents our current known understanding of the cognitive process, which in essence, is very little. Understanding that improving this model is of paramount importance, we made the decision to gather as much information about our participants as possible to establish which factors may influence the cognitive processing. Demographic features including age, sex, medical training, GP experience and a 300-point personality assessment using a validated tool from the International Personality Item Pool open source repository (IPIP), based on Costa and McRae's work [34], [35].

Being acutely aware of the risks of software-based personality testing, it was imperative to choose a valid tool [36]. The 300-point personality assessment tool was chosen to allow profiling participants in the 30 facets of personality as well as the established 'Big 5' personality domains. Previous research has highlighted that certain personality traits may be more likely to choose certain specialities [37], [38] or be more at risk of error [39]. With these findings in mind, we made the decision to ensure that our personality assessment would not lack the required granularity to sufficiently differentiate between participants or fully identify traits contributing to error.

Modelling the patient takes the form of a stream of information at a low level; symptoms, concerns, previous history. As the system evolves, GP Benchmark becomes more complex. To achieve a higher fidelity system and model the increasingly complex dynamics, GP Benchmark must assimilate systemic information in the form of declarative and procedural knowledge. Graph modelling of environmental factors, medical and procedural data becomes a further requirement in the evolution of GP Benchmark.

Due to potential skills fade [40], simulated scenarios require assessment by General Practitioners who are actively or within

6 months of practicing. Calculating a sample size based on a predicted effect is challenging, but using an error rate of 15% and using a conservative value of a further 10% increase in error rate when using biased information, approximately 826 GPs are required for a 95% chance of detecting significance at the 5% level.

Initially, 6 clinical scenarios are used to establish a baseline in the GP Benchmark simulation. The clinical scenarios are simulated cases, based on common areas of clinical practice. Each GP is tested to establish a 'non-biased' test set of data, then randomly assigned to cases developed to expose cognitive biases. Given the initial simulated environment is low fidelity, the level of GP workload will be lower than they might experience in normal operational conditions, reducing the confounding factors such as cognitive load, context-switching and time constraints experienced in normal practice. As model insights are gained, the fidelity of the environment will increase to include more complex elements, such as patient interaction, time constraints and multi-tasking.

Iterative development of the platform has enabled a detailed understanding of the domain space. We utilised insights from a small group of GPs to maintain a solution that is relevant, fit for purpose and fulfils the requirements of key stakeholders.

A significant requirement highlighted through our discussions is to develop standardised case scenarios that test cognitive bias in General Practice, as processes for developing these are currently lacking. The proposed solution is a consultation group, with expertise in clinical practice, medical education and psychology, tasked with developing a framework to create clinical scenarios from data of medical defence organisations and previous SEAs.

Additional requirements and those requiring further development include:

- a model to simulate the complex environment of GP surgery – replicating time, interruptions, tiredness, etc
- a framework for measuring and identifying type of bias in the context of clinical decision making
- a strategy to validate the insights from GP Benchmark

V. Discussion & Future Work

In this section, we discuss several areas including benefits, challenges, risks, and future work. Initial requirements introduced in this paper are focused on our initial findings, discussion with GP stakeholders and deficiencies of understanding in this research area. GP Benchmark could be used to formalise these using a requirement engineering process.

The ultimate success or failure of GP Benchmark will be with regards to its validity and generalisability to real-world clinical practice. A significant body of future work involves the project GP-VEE (General Practice Virtual Eyes & Ears), a project targeted at clinical consultation capture and generation of data to be used in the validation of insights from GP Benchmark. GP-VEE and GP Benchmark have been included in a multi-national, multi-institutional Horizon 2020 application.

Better understanding of the factors likely to increase clinical error has obvious benefits with regards to targeted mitigation. A system that can do this in real-time removes the problems with reflective interpretation and recall bias, but there is a significant recurring question regarding how certain we can be that the biases we assess are actual biases or significant to real-world clinical practice. Significance to real-world practice should be covered with the partner project GP-VEE. Validating our findings will be confirmed through reproduction and confirming the presence of a 'factor' potentially contributing to error. There are established frameworks for categorising cognitive biases, but we are also aware that with greater understanding of the domain we may need to develop a new error ontology.

We limited the scope of GP Benchmark to GP users. However, initial discussions with other health professionals, such as Advanced Care Practitioners (ACPs) highlighted their desire to be included in its development. ACPs represent a rapidly increasing workforce in General Practice and we completely acknowledge their importance as a stakeholder in GP Benchmark's development. Future work will include engagement with additional stakeholders in General Practice, expanding GP Benchmark to accommodate ACPs and other groups in its development.

Better understanding errors in clinical decision-making, particularly cognitive biases and personality, may result in unintended consequences. A potential for personality testing to be included as part of medical training, dictating career paths or recruitment. Also, publicly highlighting deficiencies in decision-making has the potential to undermine the individual and the profession as a whole. Great care in our approach to this area and clear presentation of our research should prevent inappropriate conclusions being drawn.

As a counter to the potential risks, a better understanding and appreciation of the multi-factorial nature of error may facilitate a move away from blaming the individual to a holistic approach to preventing future occurrence. Improvement in working conditions may be achieved if time, workload, or other system factors are found to be particularly culpable in error. Mindful of potential risks and benefits, GP data, such as case responses, will be private to other GPs and extracted from the platform for analysis in an aggregate form. We await confirmation of ethical approval in all aspects of our work.

VI. Conclusions

This paper highlights the initial requirements and path we have chosen in exploring the complexities and multiple factors that contribute to medical errors. These requirements include the understanding that the details and dynamics of both human cognitive process and the structure of the environment in which that process operates must be considered jointly, not in isolation from one another. GP Benchmark is our solution to understanding this complex and poorly investigated area. Initially it is a tool for better understanding the problem domain whilst also identifying future requirements.

The ultimate aim of GP Benchmark is to better model the clinical consultation so higher-fidelity simulations can be developed, mirroring clinical practice and providing a validated simulation environment. Developed in this way, GP Benchmark could have significant implications in new product testing, providing a protected domain to implement and evaluate interventions designed to mitigate error.

REFERENCES

[1] H. Singh, G. D. Schiff, M. L. Graber, I. Onakpoya, and M. J. Thompson, "The global burden of diagnostic errors in primary care," *BMJ Qual Saf*, vol. 26, no. 6, pp. 484–494, 2017.

[2] A. D. Waterman, J. Garbutt, E. Hazel, W. C. Dunagan, W. Levinson, V. J. Fraser, and T. H. Gallagher, "The emotional impact of medical errors on practicing physicians in the united states and canada," *Jt Comm J Qual Patient Saf*, vol. 33, no. 8, pp. 467–476, 2007.

[3] A. Cooper and A. Chuter, "Patient safety research in primary care: where are we now?" *Br J Gen Pract*, vol. 65, no. 641, pp. 622–623, 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4655710/

[4] K. G. Sweeney and R. Mannion, "Complexity and clinical governance: using the insights to develop the strategy." *Br J Gen Pract*, vol. 52, pp. S4–S9, 2002.

[5] O. Kostopoulou, M. Sirota, T. Round, S. Samaranayaka, and B. C. Delaney, "The role of physicians' first impressions in the diagnosis of possible cancers without alarm symptoms," *Med Decis Making*, vol. 37, no. 1, pp. 9–16, 2017.

[6] A. Huntley, D. Lasserson, L. Wye, R. Morris, K. Checkland, H. England, C. Salisbury, and S. Purdy, "Which features of primary care affect unscheduled secondary care use? a systematic review," *BMJ Open*, vol. 4, no. 5, p. e004746, 2014.

[7] N. Summerton, "Making a diagnosis in primary care: symptoms and context," *Br J Gen Pract*, vol. 54, no. 505, pp. 570–571, 2004.

[8] M. L. Graber, N. Franklin, and R. Gordon, "Diagnostic error in internal medicine," *Arch. Intern. Med.*, vol. 165, no. 13, pp. 1493–1499, 2005.

[9] E. D. O'Sullivan and S. J. Schofield, "A cognitive forcing tool to mitigate cognitive bias – a randomised control trial," *BMC Medical Education*, vol. 19, no. 1, p. 12, 2019.

[10] C. R. Goyder, C. H. Jones, C. J. Heneghan, and M. J. Thompson, "Missed opportunities for diagnosis: lessons learned from diagnostic errors in primary care," *Br J Gen Pract*, vol. 65, no. 641, pp. e838–e844, 2015.

[11] N. S. Weingart, R. M. Wilson, R. W. Gibberd, and B. Harrison, "Epidemiology of medical error," *BMJ*, vol. 320, no. 7237, pp. 774–777, 2000.

[12] J. W. Ely, L. C. Kaldjian, and D. M. D'Alessandro, "Diagnostic errors in primary care: lessons learned," *J Am Board Fam Med*, vol. 25, no. 1, pp. 87–97, 2012.

[13] B. Winters, J. Custer, S. M. Galvagno, E. Colantuoni, S. G. Kapoor, H. Lee, V. Goode, K. Robinson, A. Nakhasi, P. Pronovost, and D. Newman-Toker, "Diagnostic errors in the intensive care unit: a systematic review of autopsy studies," *BMJ Qual Saf*, vol. 21, no. 11, pp. 894–902, 2012.

[14] E. O'Sullivan and S. Schofield, "Cognitive bias in clinical medicine," *J R Coll Physicians Edinb*, vol. 48, no. 3, pp. 225–232, 2018.

[15] P. Croskerry, "Achieving quality in clinical decision making: cognitive strategies and detection of bias," *Acad Emerg Med*, vol. 9, no. 11, pp. 1184–1204, 2002.

[16] S. Coderre, H. Mandin, P. H. Harasym, and G. H. Fick, "Diagnostic reasoning strategies and diagnostic success," *Med. Educ.*, vol. 37, no. 8, pp. 695–703, 2003.

[17] Q. Cassam, "Diagnostic error, overconfidence and self-knowledge," *Palgrave Commun*, vol. 3, no. 1, pp. 1–8, 2017.

[18] P. Croskerry, "From mindless to mindful practice — cognitive bias and clinical decision making," *N Engl J Med*, vol. 368, no. 26, pp. 2445–2448, 2013.

[19] Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, and The National Academies of Sciences, Engineering, and Medicine, *Improving Diagnosis in Health Care*. National Academies Press (US), 2015.

[20] M. L. Graber, S. Kissam, V. L. Payne, A. N. D. Meyer, A. Sorensen, N. Lenfestey, E. Tant, K. Henriksen, K. LaBresh, and H. Singh, "Cognitive interventions to reduce diagnostic error: a narrative review," *BMJ Qual Saf*, vol. 21, no. 7, pp. 535–557, 2012.

[21] G. Altabbaa, A. D. Raven, and J. Laberge, "A simulation-based approach to training in heuristic clinical decision-making," *Diagnosis*, vol. 6, no. 2, pp. 91–99, 2019.

[22] O. Kostopoulou, B. C. Delaney, and C. W. Munro, "Diagnostic difficulty and error in primary care—a systematic review," *Fam Pract*, vol. 25, no. 6, pp. 400–413, 2008.

[23] H. Singh and D. F. Sittig, "Setting the record straight on measuring diagnostic errors. reply to: 'bad assumptions on primary care diagnostic errors' by Dr Richard Young," *BMJ Qual Saf*, vol. 24, no. 5, pp. 345–348, 2015.

[24] J. I. Westbrook, M. Z. Raban, S. R. Walter, and H. Douglas, "Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study," *BMJ Qual Saf*, vol. 27, no. 8, pp. 655–663, 2018.

[25] S. N. Weingart, M. Toth, D. Z. Sands, M. D. Aronson, R. B. Davis, and R. S. Phillips, "Physicians' decisions to override computerized drug alerts in primary care," *Arch. Intern. Med.*, vol. 163, no. 21, pp. 2625–2631, 2003.

[26] D. W. Bates and A. A. Gawande, "Improving safety with information technology," *N. Engl. J. Med.*, vol. 348, no. 25, pp. 2526–2534, 2003.

[27] S. S. Panesar, D. deSilva, A. Carson-Stevens, K. M. Cresswell, S. A. Salvilla, S. P. Slight, S. Javad, G. Netuveli, I. Larizgoitia, L. J. Donaldson, D. W. Bates, and A. Sheikh, "How safe is primary care? a systematic review," *BMJ Qual Saf*, vol. 25, no. 7, pp. 544–553, 2016.

[28] R. A. Elliott, E. Camacho, F. Campbell, D. Jankovic, M. S. James, E. Kaltenthaler, R. Wong, M. J. Sculpher, and R. Faria, "Prevalence and economic burden of medication errors in the NHS in england," *EEPRU*, p. 174, 2018.

[29] P. Kumarapeli and S. de Lusignan, "Using the computer in the clinical consultation; setting the stage, reviewing, recording, and taking actions: multi-channel video study," *J Am Med Inform Assoc*, vol. 20, pp. e67–e75, 2013.

[30] H. Singh, M. L. Graber, S. M. Kissam, A. V. Sorensen, N. F. Lenfestey, E. M. Tant, K. Henriksen, and K. A. LaBresh, "System related interventions to reduce diagnostic error: A narrative review," *BMJ Qual Saf*, vol. 21, no. 2, pp. 160–170, 2012.

[31] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: benefits, risks, and strategies for success," *npj Digital Medicine*, vol. 3, no. 1, pp. 1–10, 2020, number: 1 Publisher: Nature Publishing Group.

[32] J. G. Klein, "Five pitfalls in decisions about diagnosis and prescribing," *BMJ*, vol. 330, no. 7494, pp. 781–783, 2005.

[33] M. Smith, J. Higgs, and E. Ellis, "Factors influencing clinical decision making," *Clinical reasoning in the health professions*, pp. 89–100, 2008.

[34] P. T. Costa Jr. and R. R. McCrae, "Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory," *J Pers Assess.*, vol. 64, no. 1, pp. 21–50, 1995.

[35] Multi-construct IPIP inventories. [Online]. Available: https://ipip.ori.org/newMultipleconstructs.htm

[36] D. Graziotin, P. Lenberg, R. Feldt, and S. Wagner, "Behavioral software engineering: Methodological introduction to psychometrics," *arXiv:2005.09959 [cs]*, 2020. [Online]. Available: http://arxiv.org/abs/2005.09959

[37] S. Mullola, C. Hakulinen, J. Presseau, D. Gimeno Ruiz de Porras, M. Jokela, T. Hintsa, and M. Elovainio, "Personality traits and career choices among physicians in finland: employment sector, clinical patient contact, specialty and change of specialty," *BMC Med Educ*, vol. 18, no. 1, p. 52, 2018.

[38] N. J. Borges and M. L. Savickas, "Personality and medical specialty choice: A literature review and integration," *J. Career Assess*, pp. 362–380, 2002.

[39] M. Babaei, M. Mohammadian, M. Abdollahi, and A. Hatami, "Relationship between big five personality factors, problem solving and medical errors," *Heliyon*, vol. 4, no. 9, p. e00789, 2018.

[40] J. L. Oates, *Skills fade: a review of the evidence that clinical and professional skills fade during time out of practice, and of how skills fade may be measured or remediated.* General Medical Council, 2014.