Automation Bias in Electronic Prescribing

The effects of over-reliance on clinical decision support in relation to errors, cognitive load and verification.

David A. Lyell

BA., MBA.

Centre for Health Informatics

Australian Institute of Health Innovation

Faculty of Medicine and Health Sciences

MACQUARIE University

SYDNEY · AUSTRALIA

December 2018

Thesis presented for the degree of Doctor of Philosophy

Table of contents

Abstract

Background: Prescribing errors are a leading preventable cause of patient harm. Clinical decision support (CDS) can improve safety by alerting clinicians to potential errors as they enter orders into e-prescribing systems. However, this can introduce the risk of *automation bias*; clinicians may over-rely on CDS, thereby reducing vigilance in information seeking and processing. Problematically, CDS may not detect all significant errors or may generate alerts which are not clinically significant. *Omission errors* occur when clinicians fail to detect prescribing errors because they were not alerted, and *commission errors* occur where incorrect advice is wrongly acted upon. To date, there has been little research on automation bias in healthcare, where tasks, decision support and task complexity are likely to differ from those utilised in existing research which comes mostly from the heavily automated domains of aviation and process control.

This thesis examines the risk of automation bias in e-prescribing that is assisted by CDS and whether this risk is mediated by task complexity. It also examines the relationship between automation bias errors, cognitive load, and verification of CDS.

Methods: One hundred and twenty students in the final two years of a medical degree prescribed medicines for nine clinical scenarios using a simulated e-prescribing system in a randomised controlled experiment. The quality of CDS (correct, incorrect and no CDS) and task complexity (low, low with interruption and high) were varied within-subjects. *Omission errors* (failure to detect prescribing errors), *commission errors* (acceptance of false positive alerts), cognitive load, and verification of CDS (access of drug references) were measured.

Results: *Errors.* Compared to no CDS, incorrect CDS significantly increased omission errors by 33.3% ($p < .0001$), 24.5% ($p = .009$), and 26.7% ($p < .0001$) and commission errors by 65.8% ($p < .0001$), 53.5% ($p < .0001$), and 51.7% ($p < .0001$), for low-, low- with interruption and high-complexity scenarios, respectively. Task complexity and interruptions did not affect errors.

Cognitive Load. The use of CDS reduced cognitive load in high complexity conditions compared to no CDS, $F(2,117)=4.72, p=.015$. Omission errors were associated with significantly lower cognitive load with incorrect and no CDS, $F(1,636.49)=3.79, p=.023$.

Verification. Lower view times (as a percentage of task time) increased omission errors, $F(3, 361.914)=4.498$, $p=.035$, and commission errors, $F(1, 346.223)=2.712$, $p=.045$. View times were lower in CDS-assisted compared to unassisted conditions, $F(2, 335.743)=10.443$, $p<.001$.

Conclusions: This thesis contributes the first evidence of automation bias in e-prescribing, a common clinical decision-making task aided by a frequently encountered form of CDS. It also contributes the first evidence of the relationship between automation bias and reduced allocation of cognitive

resources. Participants made omission errors by failing to detect prescribing errors not alerted by CDS and made commission errors by accepting incorrect false-positive alerts. The presence of CDS reduced cognitive load and verification, and increased errors when CDS was incorrect. These effects were exacerbated under conditions of high task complexity, suggesting high complexity may be a risk factor. Curiously, however, task complexity had no effect on errors.

Participants who made automation bias errors allocated fewer cognitive resources and verified less than those who avoided errors. These findings support the cognitive miser hypothesis of automation bias that CDS alerts were used as a heuristic or mental shortcut for detecting and avoiding prescribing errors. It is highly likely that when clinicians suffer an automation bias, they reduce both verification behaviours and the cognitive resources allocated to processing information. This, in turn, compromises their ability to detect problems, which could potentially result in patient harm.

The challenge is to foster appropriate reliance on CDS, which improves efficiency and reduces errors when correct but can lead to automation bias errors when incorrect. Verification of CDS provides a key means to discriminate correct from incorrect CDS that could prevent automation bias errors. More research will be needed on how to best assist clinicians with this crucial task whilst simultaneously leveraging the enhanced efficiency and safety offered by correct CDS. Clinicians should be mindful of the limitations of CDS and the possibility that it can fail. They should be ever-vigilant and ready to verify whenever unfamiliarity or uncertainty is present, or a risk of patient harm is suspected.

Statement of originality and declarations

This work has not previously been submitted for a degree or diploma at any other university or institution. The content of this thesis represents my own original work and to the best of my knowledge and belief, contains no material previously published or written by another person except where due reference or acknowledgement is made.

The experimental research presented complied with the National Statement on Ethical Conduct in Human Research,[1] and was conducted in accordance with protocols approved by the Macquarie University Human Research Ethics Committee (Ref: 5201401029) and the University of New South Wales Medical and Community Human Research Ethics Advisory Panel (Ref: 2014-7-32).

_____ Date: _____

David A Lyell

1. National Health and Medical Research Council, Australian Research Council, Australian Vice-Chancellors' Committee. National Statement on Ethical Conduct in Human Research. *Council NHaMR* 2007 (Updated 2015) www.nhmrc.gov.au/guidelines/publications/e72

x

Acknowledgements

This research was made possible by the contributions, support and encouragement of those who generously shared their time, knowledge and expertise with me in this undertaking.

I wish to express my heartfelt thanks to my supervisors, Enrico Coiera and Farah Magrabi, for their guidance, supervision and unwavering support. I am especially thankful for the knowledge, experience and wisdom Enrico and Farah have shared with me, as well as their patience, encouragement and belief in me and this research. They shared with me their passion and vision for health informatics in providing safe, effective and accessible healthcare, and the role of scientific research in enabling this. These are values that I will bring into future endeavours. I would also like to thank Geoff McDonnell for his supervision of my initial research topic and for sharing his passion for health system simulation and systems thinking.

I wish to acknowledge the generous support of the HCF Research Foundation who provided the doctoral scholarship which afforded me the opportunity to undertake this research.

The e-prescribing experiment was made possible with the contributions of Magda Raban who designed the clinical prescribing scenarios. Ric Day advised on the design of the experimental task and provided independent review of the clinical scenarios. Lisa Pont designed the interruption task and assessed the severity of errors in the clinical scenarios and made by participants. Robin Butterfield independently reviewed and provided additional feedback on the clinical scenarios. Melissa Baysari provided advice on medication alerts and electronic medication management systems. Melissa also provided feedback on the design of the e-prescribing system. Magda, Lisa, Melissa and Ric co-authored Article II, providing valuable feedback on manuscript drafts. Monish Maharaj pilot tested the experiment and provided advice on participant recruitment.

I wish to thank Vitaliy Kim for his enormous contribution to the development of the simulated e-prescribing system used in the experiment. This was a significant undertaking, and I've learned a great deal from Vitality in the process. Jingbo Liu provided additional programming for the experiment.

John Sweller provided advice on the application of Cognitive Load Theory to this research and provided feedback and advice on the adaptation of the cognitive load inventory for use in the e-prescribing experiment. Ouhao Chen provided feedback and advice on Cognitive Load Theory and the analysis of results. Peter Petocz advised on statistical analysis; his advice to use multilevel modelling enabled the comparison between participants by whether an error was made. Thierry Wendling advised on the selection of predictors for multilevel models.

I am extremely grateful for these generous contributions, which helped me to design and conduct an experiment which has produced new insights into the operation of automation bias. I have learnt a great deal from these collaborations.

I wish to thank friends and colleagues at the Centre for Health Informatics and the Australian Institute of Health Innovation. Special thanks to Denise Tsiros who has been an amazing source of support throughout my candidature.

Finally, I would like to thank Rhonda Siu for her love and incredible support, which alone is amazingly noteworthy. However, in addition to that, Rhonda independently screened the 81 articles for inclusion in the systematic review and then independently rated the verification complexity for each of the 40 included studies. She has proof-read every manuscript and every thesis chapter. She has listened to practice runs of presentations and provided invaluable feedback, advice and support. I appreciate Rhonda is a philosopher and that this has taken her well outside her field.

List of abbreviations

AB	Automation bias
ADC	Automated dispensing cabinet
CAD	Computer-aided detection (in radiology)
CDS	Clinical decision support
ECG or EKG	Electrocardiogram
e-prescribing	Electronic prescribing
GP	General practitioner
HIT	Health information technology
NASA-TLX	NASA Task Load Index

List of original publications

Article I: Automation bias and verification complexity: A systematic review

Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association* 2017;**24**(2):423-31 doi: 10.1093/jamia/ocw105

Article II: Automation bias in electronic prescribing

Lyell D, Magrabi F, Raban MZ, Pont LG, Baysari MT, Day RO, Coiera E. Automation bias in electronic prescribing. *BMC Medical Informatics and Decision-making* 2017;**17**(1):28 doi: 10.1186/s12911-017-0425-5

Article III: The effect of cognitive load and task complexity on automation bias in electronic prescribing

Lyell D, Magrabi F, Coiera E. The effect of cognitive load and task complexity on automation bias in electronic prescribing. *Human Factors* 2018;**60**(7):1008 - 21 doi: 10.1177/0018720818781224

Article IV: Reduced verification of medication alerts increases prescribing errors

Lyell D, Magrabi F, Coiera E. Reduced verification of medication alerts increases prescribing errors. *Applied Clinical Informatics* 2019;**10**(01):066-76 doi: 10.1055/s-0038-1677009

Thesis by publication and contributions of co-authors

This thesis is presented as a series of four publications, which together form an integrated piece of research examining automation bias in healthcare that seeks to address the thesis aims set out in chapter 1.

David Lyell (DL; the candidate) conceived of, led, designed, conducted the research and made the largest contribution to each publication. DL first authored and drafted all publications, with feedback from co-authors, reviewers and editors. DL ran the experiment and conducted all analyses.

An authorship statement, specifying the contribution of all authors, is provided in the chapter introduction for each publication.

By signing below, co-authors provide their permission to include co-authored publications in this thesis and indicate their agreement with the description of authorship provided here and accompanying each co-authored article.

Enrico Coiera (EC)	Articles I, II, III & IV
Farah Magrabi (FM)	Articles II, III & IV
Magda Z. Raban (MZR)	Article II
L.G. Pont (LGP)	Article II
Melissa T. Baysari (MTB)	Article II
Richard O. Day (ROD)	Article II

1 Introduction

1.1 Background

Health information technology (HIT) has the potential to improve the quality, efficiency and safety of healthcare.[1-5] A commonly-encountered technology in healthcare is clinical decision support (CDS),[1] which provides clinicians with "knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care."[6] Decision support provided in electronic prescribing software can help to prevent adverse events by triggering medication alerts which warn clinicians of potential prescribing errors such as adverse drug interactions.[3-5]

CDS is a form of automation, where computer software perform tasks otherwise done by the clinician.[7] One such task is the application of a drug knowledge base to determine whether medications adversely interact with, or are contraindicated by, patient allergies or comorbidities. This can assist clinicians in making prescribing decisions but does not prevent them from applying their own knowledge, judgement and seeking further information in addition to that provided by CDS. It is the clinician who decides what treatment to prescribe and bears ultimate responsibility for the outcomes of those treatment decisions.

Clinical judgement and oversight are essential, especially as CDS is imperfect and can be incorrect. CDS can malfunction, causing alerts to trigger in situations when they should not, failing to trigger alerts when they should, and displaying alerts that suggest the wrong action.[8, 9] Malfunctions can occur due to programming errors in the CDS software, the incorrect conceptualisation of rules triggering alerts and mismatches between clinical problems and data sources used by CDS.[8, 9] The quality of patient-specific alerts is dependent on the accuracy and completeness of patient records. Missing or inaccurate information can lead to incorrect and unsafe advice.[10] Likewise, data entered improperly or in the wrong field cannot be used by CDS.[11] Finally, there are marked variations between the systems provided by different vendors regarding the types of errors detected and alerted by CDS.[12] This could lead to situations where clinicians form incorrect assumptions about the capabilities of a system based on prior experience with another system from a different vendor.

1.2 Automation bias

Over-reliance on CDS can lead to errors with the potential for patient harm when CDS alerts are incorrect. This over-reliance is known as *automation bias* (AB), which Mosier and Skitka [13] define as "the tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing." Automation bias can lead to two different types of errors. *Omission errors* occur when

people fail to notice problems because they were not alerted to them by decision support; they involve a failure to act. *Commission errors* occur when people comply with incorrect decision support recommendations; they involve wrong actions.[14]

Importantly, the classification of errors as omission or commission relates directly to the task that is assisted by decision support. This work focuses on CDS which automates the task of detecting prescribing errors by triggering medication alerts. For example, a clinician would make an omission error if they fail to detect a prescribing error because they were not alerted to it by CDS (a CDS false-negative); this would result in a failure to act to avoid the error. A commission error occurs when a clinician performs a wrong action by agreeing with incorrect CDS advice (a CDS false-positive). For example, a clinician might be discouraged from prescribing a gold standard treatment due to an incorrect alert.

There is also a closely-related field of research which describes *automation-induced complacency*,[15] where complacency is the "self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state."[16] Parasuraman and Manzey [17] suggest that automation bias and automation-induced complacency are overlapping manifestations of over-reliance on automation, which Parasuraman and Riley [7] refer to as automation *misuse*. Automation bias omission errors and automation-induced complacency both manifest as the failure of the user to detect problems because they were not alerted by automation. However, Parasuraman and Manzey [17] also suggest that automation bias commission errors display complacency-like effects, especially in relation to the verification of automated recommendations.

The focus of this work is over-reliance on decision support that leads to error, a phenomenon described by both automation bias and automation-induced complacency. Accordingly, this thesis incorporates both bodies of literature and, for the sake of brevity, will use the term, "automation bias". The classifications of omission and commission errors [13, 14] provide a useful distinction between errors made by clinicians resulting from CDS false-negatives and false-positives.

1.3 Automation bias in healthcare

There has been little study of automation bias in healthcare.[18, 19] However, there has been a number of documented incidents describing automation bias errors. These incidents provide examples of how clinical automation and decision support can be incorrect and can thereby adversely influence the users of these technologies. They also demonstrate the risk of patient harm which can follow from automation bias, thereby also providing a rationale for conducting further study of this phenomenon.

Omission errors

1. A physician who prescribed contraindicated hypertensives to a pregnant woman, resulting in the death of her unborn baby, put the error down to the lack of a 'red flag' warning from the

computer system. The inquiry report noted that CDS alerts were not activated at the time of the incident,[20] of which the physician seemed unaware. This incident occurred because the physician failed to detect a prescribing error that was not alerted by CDS.

2. A patient with usually well-controlled type 1 diabetes suffered persistent hyperglycaemia (high blood sugar levels) upon returning from an overseas trip. He consulted with doctors to stabilise his blood sugar but instead ended up being hospitalised with diabetic ketoacidosis. His insulin pump had failed, possibly from exposure to airport security scanners. The failure was silent; there was no warning or indication that the pump was not working.[21] This likely contributed to the delay in identifying and rectifying the cause of the hyperglycaemia. There are a multitude of issues which can cause hyperglycaemia. However, failure of the insulin pump was not considered until the patient was hospitalised, delaying the initiation of appropriate treatment.

3. A consultant paediatrician made a complaint about not being informed by telephone of an abnormally high conjugated bilirubin blood test result, a possible indicator of liver disease. A bug in the laboratory's software meant that no reference range was applied to the test result and, consequently, the result was not flagged as high, and the consultant was not informed. A root cause analysis noted that nine different people saw the test result, but none recognised it as out of range.[22] Staff missed the high result because of the absence of a high result flag. The root cause analysis report and permission to describe this incident were provided on the condition that it be reported in a de-identified manner.

Commission errors

4. An elderly patient who was admitted to hospital suffering new-onset seizures was given the wrong medication after pharmacy staff mis-entered the medicine into the computer system as *diltiazem* (a cardiac drug) instead of *dilantin* (used to treat seizures). The nurse responsible for administering medications recognised the discrepancy between the medication administration record which listed the correct prescription of *dilantin* and the automated dispensing cabinet (ADC) which listed the incorrectly dispensed *diltiazem*. However, the nurse relied on the information from the ADC and administered the incorrect medication, which resulted in the patient experiencing significant side effects.[23] The nurse wrongly acted on the ADC over the correct, handwritten medical record.

5. A retrospective study examined all electrocardiograms (ECG) from a metropolitan US hospital with a computer interpreted diagnosis of atrial fibrillation over a six-month period (n = 1085 patients). Each ECG was reviewed by two independent and blinded electrophysiologists. The computerised diagnosis of atrial fibrillation was incorrect for 35% of patients (n = 382). Ordering physicians failed to correct the misdiagnosis for 24% of those cases (n = 92), which

led to unnecessary changes to treatment for 10% of patients (n = 39). Two patients developed complications as a result of unnecessary changes to their treatment.[24] The changes in treatment resulting from the incorrect computerised diagnoses are commission errors.

6. During surgery, an automated blood pressure monitor, which inflates at specified intervals to measure the patient's blood pressure, showed the patient as having hypertension. The anaesthetists began treating the condition and set the monitor to update more frequently. Unfortunately, they failed to press the 'start' button to reinitiate the device, meaning that no further measurements were taken. Instead, the device continued to display the last taken measurement which indicated a state of hypertension. For the next 45 minutes, anaesthetists aggressively treated the patient for hypertension with powerful drugs, until it was discovered that the blood pressure reading displayed was not current, by which time the patient's blood pressure had been over-corrected and was very low.[25] In this incident, anaesthetists relied on the monitor, which, as Gaba [25] notes, had a very significant design flaw, to provide feedback on the effects of their interventions on the patient's blood pressure. If they had had accurate measurements, they would have treated less aggressively, ceasing once the condition had stabilised.

There are a small number of experimental studies which have directly examined over-reliance on decision support in healthcare applications.

7. A series of studies examined the computer-aided detection (CAD) of cancers in screening mammography. CAD assists readers by placing prompts on features which may indicate the presence of a cancer.[26] CAD has good sensitivity [27] but low specificity,[26] providing a large number of false-positive prompts.[26, 28] This leads to concerns that the use of CAD may result in an increase in patients being subjected to unnecessary and invasive follow-up testing.[29] Several studies have assessed this by having participants read mammograms with and without CAD assistance, and then deciding whether to recall the case for follow-up testing. There has been no evidence of AB commission errors, that is, readers recalling cases because of false positive prompts.[29] However, readers detected 25% fewer cancers that were not prompted by CAD, compared to the same cases when read unaided.[30] These are omission errors. While prompts were not relied on to detect cancers, their absence was taken as a strong indication that there was no cancer present, leading Alberdi, et al. [30] to suggest that "in many instances, the absence of prompts is more informative than their presence."

8. Golchin and Roudsari [31] asked general practitioners (GPs) to answer questions about clinical scenarios. For difficult scenarios, GPs made significantly more errors when provided with incorrect decision support. These are AB commission errors.

9. Goddard, et al. [32] asked GPs what they would prescribe for clinical scenarios, after which they were shown simulated decision support advice, some of which was incorrect. Participants were then asked if they wished to change their responses. When decision support advice was incorrect, GPs were more likely to switch a correct decision to an incorrect one. These are AB commission errors.

These incidents and studies demonstrate how: (1) CDS can produce incorrect advice; (2) incorrect advice can adversely influence decisions; and (3) these decisions can lead to patient harm. This thesis is concerned with the second aspect; it examines the interaction between clinician and incorrect CDS, that is, the point where errors can be either avoided or propagated. Automation bias can adversely alter outcomes. Further study in this area is important for ensuring patient safety as new decision support tools are constantly introduced.

1.4 Automation bias effects

Mosier and Skitka's [13] definition of automation bias characterises it as a heuristic, where over-reliance stems from the use of automation as a mental shortcut in place of the user's own efforts in information seeking and processing. This is the *cognitive miser* hypothesis of automation bias,[13] which suggests that people are misers when it comes to utilising their cognitive resources, preferring to seek adequate, faster and less effortful ways of thinking, rather than engaging in more accurate, but slower and more effortful thinking.[33] This is achieved through the use of heuristics or mental shortcuts, such as relying on CDS to identify errors, instead of reading monographs and evaluating the information to determine whether a drug therapy is safe. This hypothesis is illustrated in Figure 1-1, where an increase in automation bias is expected to increase errors, but decrease information seeking and information processing.

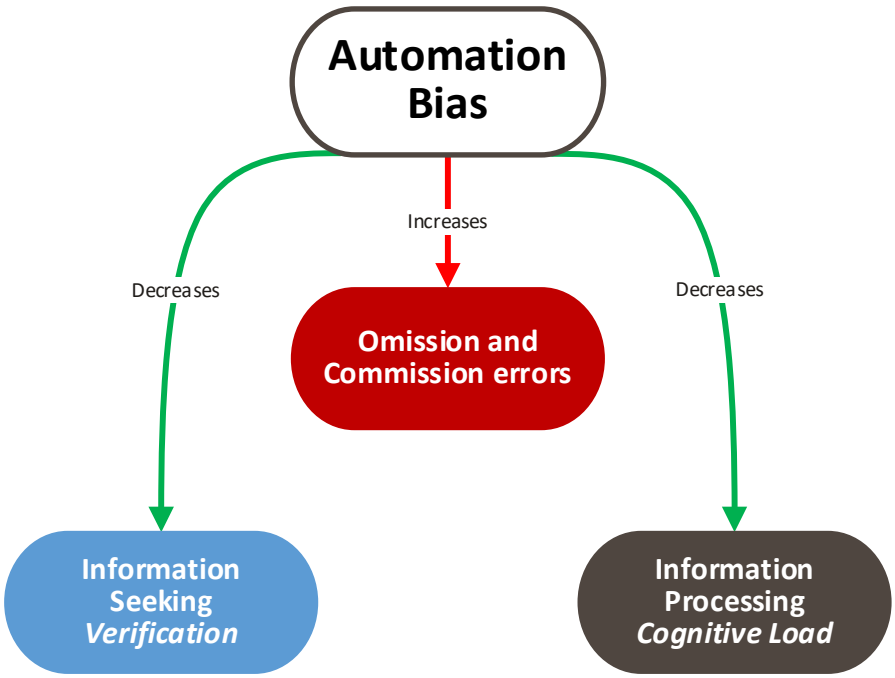

Figure 1-1 The hypothesised effect of automation bias on errors, information seeking and processing

Figure 1-1 shows the three observable outcomes of automation bias that will be examined: errors, information processing and information seeking.

1.4.1 Errors

The effect of automation bias is established through the measurement of omission and commission errors showing that incorrect CDS increases errors above that which would have occurred with no CDS.

Omission errors are usually tested by having people detect specific events, such as whether a mammogram shows the presence of a cancer [30] or if there is a problem with an aircraft's engines.[15] In these tasks, automation assists by alerting the user to the problem. Because no warning is given when automation fails, over-reliance on it produces the risk of problems going undiscovered.

To date, most experimental research has established the presence of omission errors by observing the difference between: (1) constant high accuracy automation which has consistently produced higher rates of omission errors and (2) automation that varied between high and low accuracy.[15, 34-38] This effect is likely due to participants having greater trust in automation that is highly accurate, a disposition which makes them less likely to detect automation failures.[34, 39]

Omission errors have also been tested in studies which compared incorrect automation with a non-automated control. These studies showed that when automation was incorrect, errors were higher compared to when there was no automation.[30, 40-42]

Commission errors are induced by incorrect decision support advice and would not have occurred without the automated recommendation. A few studies have found evidence of commission errors in controlled experiments.[31, 32, 43, 44] Other studies have found that participants performed actions they would not have otherwise performed, for example, shutting down an aircraft engine based on an incorrect decision support recommendation, but in the absence of any other information indicating the need to do so.[14, 45]

The effect of over-reliance on, and over-compliance with, automation on errors substantiates the existence of automation bias. Automation bias has been widely studied within the human factors and ergonomics literature, especially focused on tasks which take place in highly automated environments, such as aviation and process control.

1.4.2 Information seeking / Verification

One hypothesised effect of automation bias is that it reduces information seeking. Information seeking enables *verification*, the process of establishing the truth or correctness of something through the investigation or evaluation of data.[46] Verification plays a key role in the detection of errors by enabling the user to: (1) establish whether automation is functioning correctly and (2) identify errors independent of automation.

Research has shown that reduced verification is associated with omission errors [34, 47] and commission errors.[48-52] Omission errors were associated with fewer eye gaze fixations on the relevant information, while commission errors were associated with less access of data which could confirm or invalidate decision support recommendations.

1.4.3 Information processing / Cognitive Load

Information processing is a cognitive task whereby the acquired information is evaluated, and decisions are made based on this information. For this thesis, information processing is operationalised using Cognitive Load Theory,[53] which is based on the idea that human information processing is limited by the capacity of working memory.[53] The latter, in turn, has a limited capacity [54, 55] and short duration.[56] When information processing takes place in working memory, [57] it generates *cognitive load.*[53] The rationale for this choice is presented in Article I (chapter 2) and its measurement is described in Article III (chapter 4).

There have been a number of studies which measured workload using the NASA Task Load Index (NASA-TLX),[58] which measures workload across six dimensions: mental, physical and temporal demands, frustration, effort and performance. The NASA-TLX has been extensively used in research.[59]

Prior research has found that high constant accuracy automation which induces higher rates of automation bias is associated with lower workload.[37, 44, 60] Prinzel et al.,[37, 61] divided

participants into high and low complacency-potential groups based on their scores from the Complacency-Potential Rating Scale,[62, 63] a scale designed to rate an individual's potential to be complacent towards automation. High complacency-potential participants made significantly more omission errors than low complacency-potential participants with highly-accurate automation, a condition known to increase the rate of omission errors.[15, 34-38] Both groups made the same number of errors with variable accuracy automation, although participants in the high complacency-potential group reported a significantly higher workload for the same level of performance.

To date, the relationship between automation bias and information processing remains largely unexplored. Prinzel et al.[37, 61] demonstrates that there are differences in subjective workload based on a person's complacency-potential. This finding needs to be extended further to identify whether there are differences in information processing between instances where automation bias errors are made and when they are avoided.

To date, no studies have directly compared differences in cognitive load between people who do and do not make automation bias errors within the same experimental conditions. Hence, one key focus of this research is the impact of task complexity on automation bias, in particular, whether over-reliance is more likely in situations where people are cognitively overloaded by the information processing requirements of the task.[19] While cognitive load is a likely sub-component of workload, it provides a suitable framework for the manipulation and assessment of the effects of task complexity on information processing.

1.5 Limited automation bias research in healthcare

Automation bias in healthcare is still a relatively new field of study with a sparse body of literature.[18, 19] The incidental nature of the reporting of automation bias in healthcare limits our understanding of the extent to which it occurs and the risks it poses. This is further complicated by reports which do not explicitly identify automation bias.[18] Indeed, while many of the incidents discussed in section 1.3 were described as automation bias or automation-induced complacency, it is very likely that many more incidents exist, but have not been identified.

However, there exists a substantial body of research in the human factors and ergonomics literature, the majority of which focuses on tasks derived from aviation and process control applications.[19] Automation bias incidents in aviation were identified by analysing the Aviation Safety Reporting System (ASRS), a voluntary reporting system which allows pilots, crew, air traffic controllers, and others to report incidents and near misses confidentially with the aim of improving air safety. The ASRS has operated for over 40 years and recorded over 1.4 million reports, all of which are available via a publicly accessible database on their website.[64]

While less well-established, the reporting of health information technology-related incidents is gathering pace.[65] A group of researchers from the United States has recently reported cases of a number of CDS malfunctions [8, 9] and is actively soliciting CDS malfunctions case reports from users.[9] The focus of this automation bias research concerns how such malfunctions may impact users and contribute to errors. While such reports provide valuable insights into how CDS can be incorrect and identify areas for improvement, there remains a need to establish how incorrect CDS impacts clinicians. Until a larger and more systematic body of incident reports is collected, researchers are limited to the incidental reporting of incidents, such as those described in section 1.3.

As most of the current research on automation bias comes from the human factors and ergonomics literature, it is important to determine how applicable the existing research is to healthcare. It is likely that the tasks performed, and the role of decision support in assisting people with these tasks will be different to those in aviation and process control tasks. Therefore, it will be critical to compare and contrast these two bodies of literature with a view to understanding which research is transferable from the human factors literature and which factors are unique to healthcare and require further study.

1.6 Thesis aims

This thesis seeks to study automation bias in a healthcare context, focusing on tasks, automation, and factors characteristic of healthcare settings and applications. It focuses in particular on important healthcare factors that do not feature in the established literature which is currently dominated by aviation and process control studies. A systematic review of the automation bias literature informs the aims and hypotheses of the experimental portion of the thesis. The review also informs the choice of task in the experiment.

The experimental portion of the thesis tests for the presence of automation bias in e-prescribing when assisted by CDS medication alerts. The experiment also tests whether task complexity is a cause of automation bias and if the relationship between these two factors might be explained by cognitive overload.

An e-prescribing task assisted by CDS was chosen for the experiment for two reasons. First, it represents a task and automation type typical of the healthcare studies reported in the review, that is, a decision-making task where automated assistance is provided in the form of decision support. Second, the earlier example of a physician who blamed a prescribing error on the lack of CDS alerts demonstrates that e-prescribing may be susceptible to automation bias omission errors. However, to date, there have been no studies assessing this risk. This knowledge gap thus needs to be filled, especially in light of the large volume of prescriptions ordered, the increasing prevalence of e-prescribing that involves CDS, and the risk of harm posed by avoidable prescribing errors. A

randomised controlled experiment into automation bias in electronic prescribing would fill this gap and provide a vehicle to test hypotheses concerning task complexity and cognitive load.

Accordingly, this thesis seeks to:

Aim 1 Identify key tasks, automation (or decision support), and risk factors that are likely to be unique to, or feature predominately in, automation bias within healthcare contexts.

Aim 2 Experimentally test whether there is a risk of automation bias in electronic prescribing assisted by medication alerts from clinical decision support.

Aim 3 Experimentally test whether high-complexity tasks are more susceptible to automation bias errors.

Aim 4 Experimentally test whether task interruptions increase the rate of automation bias errors.

Aim 5 Experimentally test whether automation bias errors are caused by high cognitive load brought about by high task complexity. This may indicate that reliance on decision support serves as a coping mechanism when the cognitive demands of a task overwhelm the available cognitive resources.

Aim 6 Determine the relationship between automation bias errors, cognitive load and verification.

1.7 Thesis structure

This thesis is presented as a series of four publications which constitute chapters two to five. Three articles have been published in peer-reviewed health informatics journals and one has been published in a peer-reviewed human factors journal. A brief introduction prefaces each of these chapters to orient the reader to how the publication contributes to the thesis' aims.

Chapter two presents *Article I*,[19] a systematic review comparing the healthcare and human factors literature on automation bias, with a focus on identifying the tasks, automation and risk factors that are unique to or feature predominately in, healthcare. This article seeks to address aim 1.

An electronic prescribing experiment was conducted; it was designed to answer the questions posed in aims 2 to 6. Chapters three to five report the results of this experiment for each of the automation bias effects: errors (chapter 3), cognitive load (chapter 4) and verification (chapter 5).

Chapter three presents *Article II* [66] which reports the omission and commission errors made by the participants in the experiment. It evaluates the hypotheses set out in aims 2 to 4, that is: whether (1) electronic prescribing assisted by clinical decision support is susceptible to automation bias errors; (2) high complexity tasks are more susceptible to automation bias errors; and (3) interrupted tasks are more susceptible to automation bias errors.

Chapter four presents *Article III* [67] which reports the outcome of the experiment on participants' cognitive load and evaluates the hypothesis set out in aim 5, that is, whether high cognitive load induces automation bias.

Chapter five presents *Article IV* [68] which reports the outcome of the experiment on participants' information seeking or verification behaviours.

Chapters three to five conclude by constructing a model of the effect of CDS and task complexity on the relevant dependent variables: errors (chapter 3), cognitive load (chapter 4) and verification (chapter 5).

Chapter six presents the discussion for the overall thesis as an integrated piece of research on automation bias in healthcare. It synthesises the automation bias effects reported in previous chapters to: (1) address aim 6 by exploring the relationship between errors, cognitive load and verification, and (2) evaluate the overall impact of task complexity across all effects. The contribution, implications, and limitations of this thesis, as well as recommendations for further research, are also discussed.

1.8 Chapter 1 References

1. Chaudhry B, Wang J, Wu S, et al. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine* 2006;**144**(10):742-52 doi: 10.7326/0003-4819-144-10-200605160-00125

2. Beeuwkes Buntin M, Burke MF, Hoaglin MC, Blumenthal D. The Benefits Of Health Information Technology: A Review Of The Recent Literature Shows Predominantly Positive Results. *Health Affairs* 2011;**30**(3):464-71 doi: 10.1377/hlthaff.2011.0178

3. Wolfstadt JI, Gurwitz JH, Field TS, Lee M, Kalkar S, Wu W, Rochon PA. The Effect of Computerized Physician Order Entry with Clinical Decision Support on the Rates of Adverse Drug Events: A Systematic Review. *Journal of General Internal Medicine* 2008;**23**(4):451-58 doi: 10.1007/s11606-008-0504-5

4. Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The Effect of Electronic Prescribing on Medication Errors and Adverse Drug Events: A Systematic Review. *Journal of the American Medical Informatics Association* 2008;**15**(5):585-600 doi: 10.1197/jamia.M2667

5. van Rosse F, Maat B, Rademaker CMA, van Vught AJ, Egberts ACG, Bollen CW. The Effect of Computerized Physician Order Entry on Medication Prescription Errors and Clinical Outcome in Pediatric and Intensive Care: A Systematic Review. *Pediatrics* 2009;**123**(4):1184-90 doi: 10.1542/peds.2008-1494

6. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A Roadmap for National Action on Clinical Decision Support. *Journal of the American Medical Informatics Association* 2007;**14**(2):141-45 doi: 10.1197/jamia.M2334

7. Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse. *Human Factors* 1997;**39**(2):230-53 doi: 10.1518/001872097778543886

8. Wright A, Hickman T-TT, McEvoy D, Aaron S, Ai A, Andersen JM, Hussain S, Ramoni R, Fiskio J, Sittig DF, Bates DW. Analysis of clinical decision support system malfunctions: a case series and survey. *Journal of the American Medical Informatics Association* 2016;**23**(6):1068-76 doi: 10.1093/jamia/ocw005

9. Wright A, Ai A, Ash J, Wiesen JF, Hickman T-TT, Aaron S, McEvoy D, Borkowsky S, Dissanayake PI, Embi P, Galanter W, Harper J, Kassakian SZ, Ramoni R, Schreiber R, Sirajuddin A, Bates DW, Sittig DF. Clinical decision support alert malfunctions: analysis and empirically derived taxonomy. *Journal of the American Medical Informatics Association* 2017;**25**(5):496-506 doi: 10.1093/jamia/ocx106

10. Berner ES, Kasiraman RK, Yu F, Ray MN, Houston TK. Data Quality in the Outpatient Setting: Impact on Clinical Decision Support Systems. *AMIA Annual Symposium Proceedings* 2005;**2005**:41-45

11. Campbell EM, Sittig DF, Guappone KP, Dykstra RH, Ash JS. Overdependence on Technology: An Unintended Adverse Consequence of Computerized Provider Order Entry. *AMIA Annual Symposium Proceedings* 2007;**2007**:94-98

12. Sweidan M, Williamson M, Reeve JF, Harvey K, O'Neill JA, Schattner P, Snowdon T. Evaluation of features to support safety and quality in general practice clinical software. *BMC Medical Informatics and Decision Making* 2011;**11**(1):1-8 doi: 10.1186/1472-6947-11-27

13. Mosier KL, Skitka LJ. Human decision makers and automated decision aids: Made for each other. In: Parasuraman R, Mouloua M, eds. Automation and human performance: Theory and applications. Hillsdale, NJ, England: Lawrence Erlbaum Associates, 1996:201-20.

14. Mosier KL, Skitka LJ, Heers S, Burdick M. Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology* 1998;**8**(1):47-63 doi: 10.1207/s15327108ijap0801_3

15. Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced "complacency.". *The International Journal of Aviation Psychology* 1993;**3**(1):1-23 doi: 10.1207/s15327108ijap0301_1

16. Billings C, Lauber J, Funkhouser H, Lyman G, Huff E. NASA aviation safety reporting system. (Technical Report TM-X-3445). 1976

17. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: An attentional integration. *Human Factors* 2010;**52**(3):381-410 doi: 10.1177/0018720810376055

18. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 2012;**19**(1):121-27 doi: 10.1136/amiajnl-2011-000089

19. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association* 2017;**24**(2):423-31 doi: 10.1093/jamia/ocw105

20. Professional Standards Committee. An inquiry into a complaint in relation to Dr Sunil Kumar Dan. *NSW Health Care Complaints Commission* 24 June 2016 http://www.hccc.nsw.gov.au/ArticleDocuments/246/DAN%20Sunil%20-%20Decision%20-%20PSC%20-%2010%20June%202016.pdf.aspx

21. The Naomi Berrie Diabetes Center. Airport Scanners and Insulin Pumps: A Cautionary Report. Secondary Airport Scanners and Insulin Pumps: A Cautionary Report 2013. Retrieved from http://www.nbdiabetes.org/news/airport-scanners-and-insulin-pumps-cautionary-tale [archive link http://www.webcitation.org/6uxFgkJBT]

22. Anonymous. Root Cause Analysis: Conjugated Bilirubin. n.d.

23. ISMP Canada. Understanding human over-reliance on technology. *ISMP Canada Safety Bulletin* 2016;**16**(5):1 - 4

24. Bogun F, Anh D, Kalahasty G, Wissner E, Bou Serhal C, Bazzi R, Weaver WD, Schuger C. Misdiagnosis of atrial fibrillation and its clinical consequences. *American Journal of Medicine* 2004;**117**(9):636-42 doi: 10.1016/j.amjmed.2004.06.024

25. Gaba D. Automation in anesthesiology. In: Mouloua M, Parasuraman R, eds. Human performance in automated systems: Current research and trends. Hillsdale, NJ: Lawrence Erlbaum Associates, 1994:57-63.

26. Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H. Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. *Health Technology Assessment* 2005;**9**(6):1-58

27. Murakami R, Kumita S, Tani H, Yoshida T, Sugizaki K, Kuwako T, Kiriyama T, Hakozaki K, Okazaki E, Yanagihara K, Iida S, Haga S, Tsuchiya S. Detection of breast cancer with a computer-aided detection applied to full-field digital mammography. *Journal Of Digital Imaging* 2013;**26**(4):768-73 doi: 10.1007/s10278-012-9564-5

28. Alberdi E, Povyakalo AA, Strigini L, Ayton P, Hartswood M, Procter R, Slack R. Use of computer-aided detection (CAD) tools in screening mammography: a multidisciplinary investigation. *The British Journal of Radiology* 2005;**78**(suppl_1):S31-S40 doi: doi:10.1259/bjr/37646417

29. Marx C, Malich A, Facius M, Grebenstein U, Sauner D, Pfleiderer SOR, Kaiser WA. Are unnecessary follow-up procedures induced by computer-aided diagnosis (CAD) in mammography? Comparison of mammographic diagnosis with and without use of CAD. *European Journal of Radiology* 2004;**51**(1):66-72 doi: 10.1016/S0720-048X(03)00144-X

30. Alberdi E, Povykalo A, Strigini L, Ayton P. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. *Academic Radiology* 2004;**11**(8):909-18 doi: 10.1016/j.acra.2004.05.012

31. Golchin K, Roudsari A. Study of the effects of clinical decision support system's incorrect advice and clinical case difficulty on users' decision making accuracy. *Studies in Health Technology and Informatics* 2011;**164**:13-16 doi: 10.3233/978-1-60750-709-3-13

32. Goddard K, Roudsari A, Wyatt JC. Automation bias: empirical results assessing influencing factors. *International Journal of Medical Informatics* 2014;**83**(5):368-75 doi: 10.1016/j.ijmedinf.2014.01.001

33. Fiske ST, Taylor SE. *Social cognition*. New York: Random House, 1984.

34. Bagheri N, Jamieson GA. Considering subjective trust and monitoring behavior in assessing automation-induced "complacency". In: Vincenzi DA, Mouloua M, Hancock PA, eds. Human Performance, Situation Awareness and Automation: Current Research and Trends, Vol 2. Mahwah: Lawrence Erlbaum Associates, 2004:54-59.

35. Bailey NR, Scerbo MW, Freeman FG, Mikulka PJ, Scott LA. Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors* 2006;**48**(4):693-709 doi: 10.1518/001872006779166280

36. Parasuraman R, de Visser E, Lin M-K, Greenwood PM. Dopamine beta hydroxylase genotype identifies individuals less susceptible to bias in computer-assisted decision making. *PLoS ONE* 2012;**7**(6) doi: 10.1371/journal.pone.0039675

37. Prinzel LJ, III, Freeman FG, Prinzel HD. Individual Differences in Complacency and Monitoring for Automation Failures. *Individual Differences Research* 2005;**3**(1):27-49

38. Singh IL, Singh AL, Saha PK. Monitoring performance and mental workload in an automated system. Proceedings of the International Conference onEngineering Psychology and Cognitive Ergonomics; 2007 Jul 22-27; Beijing, China. Springer Verlag.

39. Bailey NR, Scerbo MW. Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science* 2007;**8**(4):321-48 doi: 10.1080/14639220500535301

40. Singh IL, Sharma HO, Parasuraman R. Effects of manual training and automation reliability on automation induced complacency in flight simulation task. *Psychological Studies* 2001;**46**(1/2):21-27

41. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making? *International Journal of Human Computer Studies* 1999;**51**(5):991-1006 doi: 10.1006/ijhc.1999.0252

42. Metzger U, Parasuraman R. Automation in future air traffic management: effects of decision aid reliability on controller performance and mental workload. *Human Factors* 2005;**47**(1):35-49 doi: 10.1518/0018720053653802

43. Sarter NB, Schroeder B. Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors* 2001;**43**(4):573-83 doi: 10.1518/001872001775870403

44. Rovira E, McGarry K, Parasuraman R. Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors* 2007;**49**(1):76-87 doi: 10.1518/001872007779598082

45. Mosier KL, Skitka LJ, Dunbar M, McDonnell L. Aircrews and automation bias: The advantages of teamwork? *International Journal of Aviation Psychology* 2001;**11**(1):1-14 doi: 10.1207/s15327108ijap1101_1

46. Oxford English Dictionary. *"verification, n."*: Oxford University Press, June 2018.

47. Bagheri N, Jamieson GA. The impact of context-related reliability on automation failure detection and scanning behaviour. IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583); 2004 Oct 10-13.

48. Bahner J, Huper A-D, Manzey D. Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies* 2008;**66**(9):688-99 doi: 10.1016/j.ijhcs.2008.06.001

49. Bahner J, Elepfandt MF, Manzey D. Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2008 Sep 22-26; New York, NY, United states. Human Factors And Ergonomics Society.

50. Manzey D, Reichenbach J, Onnasch L. Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience. *Journal of Cognitive Engineering and Decision Making* 2012;**6**(1):57-87 doi: 10.1177/1555343411433844

51. Reichenbach J, Onnasch L, Manzey D. Misuse of automation: The impact of system experience on complacency and automation bias in interaction with automated aids. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2010 Sep 27 - Oct 1; San Francisco, CA, United states. Human Factors And Ergonomics Society.

52. Reichenbach J, Onnasch L, Manzey D. Human performance consequences of automated decision aids in states of sleep loss. *Human Factors* 2011;**53**(6):717-28 doi: 10.1177/0018720811418222

53. Sweller J, Ayres P, Kalyuga S. *Cognitive load theory*. New York: Springer, 2011.

54. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 2001;**24**:87-185

55. Miller GA. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 1956;**63**(2):81

56. Peterson L, Peterson MJ. Short-term retention of individual verbal items. *Journal of Experimental Psychology* 1959;**58**(3):193-98 doi: 10.1037/h0049234

57. Baddeley A. Working Memory. *Science* 1992;**255**(5044):556

58. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 1988;**52**:139-83

59. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 2006;**50**(9):904-08 doi: 10.1177/154193120605000909

60. Singh AL, Tiwari T, Singh IL. Effects of automation reliability and training on automation-induced complacency and perceived mental workload. *Journal of the Indian Academy of Applied Psychology* 2009;**35**(spec iss):9-22

61. Prinzel III LJ, DeVries H, Freeman FG, Mikulka P. Examination of automation-induced complacency and individual difference variates (Technical Report NASA / TM-2001-211413). *National Aeronautics and Space Administration Langley Research Center* 2001 https://ntrs.nasa.gov/search.jsp?R=20020021642

62. Singh IL, Molloy R, Parasuraman R. Automation-induced "complacency": Development of the Complacency-Potential Rating Scale. *The International Journal of Aviation Psychology* 1993;**3**(2):111-22 doi: 10.1207/s15327108ijap0302_2

63. Singh IL, Molloy R, Parasuraman R. Individual Differences in Monitoring Failures of Automation. *The Journal of General Psychology* 1993;**120**(3):357-73 doi: 10.1080/00221309.1993.9711153

64. NASA Aviation Safety Reporting System. ASRS Program Briefing. *NASA Aviation Safety Reporting System* 2016 https://asrs.arc.nasa.gov/docs/ASRS_ProgramBriefing2016.pdf

65. Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. *Journal of the American Medical Informatics Association* 2017;**24**(2):246-50 doi: 10.1093/jamia/ocw154

66. Lyell D, Magrabi F, Raban MZ, Pont LG, Baysari MT, Day RO, Coiera E. Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making* 2017;**17**(1):28 doi: 10.1186/s12911-017-0425-5

67. Lyell D, Magrabi F, Coiera E. The effect of cognitive load and task complexity on automation bias in electronic prescribing. *Human Factors* 2018;**60**(7):1008 - 21 doi: 10.1177/0018720818781224

68. Lyell D, Magrabi F, Coiera E. Reduced Verification of Medication Alerts Increases Prescribing Errors. *Applied Clinical Informatics* 2019;**10**(01):066-76 doi: 10.1055/s-0038-1677009

2 Systematic review

2.1 Background

While automation bias is a relatively new area of enquiry in healthcare,[1, 2] it is not a new field of study and there exists a substantial body of research.[1-3] Most of what is already known about automation bias comes from studies using tasks and automation found in the heavily automated domains of aviation and process control,[1, 2] where automation typically assists users with monitoring system parameters, such as aircraft engines,[4-17] or the life support system of a spacecraft.[18-22] In these domains, automation can also provide decision support to assist users with the diagnosis and resolution of system faults. The user's role is to provide supervisory control, ensuring that automation is doing what it should and to take over when things go awry.

Healthcare and clinical tasks which may be susceptible to automation bias are likely to differ from those which have been studied and tested in the existing human factors and ergonomics literature. Therefore, it is necessary to establish the extent to which the existing literature is applicable to healthcare. Specifically, there is a need to identify: (1) which factors in the existing literature can be extended to the healthcare domain, and (2) which factors are unique to healthcare and require further study.

The journal article presented in this chapter reports a systematic review comparing and contrasting the human factors and healthcare literature with a specific focus on tasks, automation, and risk factors.

2.2 Contribution of this article to the thesis

The systematic review (Article I) seeks to address Aim 1 of the thesis: to systematically review the human factors and healthcare literature by: (1) comparing and contrasting tasks, including the role of automation in those tasks, and (2) identifying unique risk factors.

The goal of this review was to identify potential gaps in the research, especially factors which are important for automation bias in healthcare that have not been studied in the existing literature.

The review found evidence of automation bias in single tasks,[2] which is contrary to the prevailing view in the human factors literature that it only occurs in multi-task environments.[7, 8] Single tasks which produced automation bias involved: (1) diagnostic tasks requiring the identification of the current state of a system or the cause of a problematic state, and (2) higher verification complexity, that is, the task complexity of verifying automation. Both of these aspects are characteristic of healthcare tasks.

These findings suggest that task complexity, and especially the complexity involved in verifying automation, may be an important risk factor for automation bias in healthcare. Task complexity is a key focus of the experimental portion of the thesis. The review concludes by proposing Cognitive Load Theory as a methodology for studying task complexity in relation to automation bias.[2]

2.3 Article details

This article was published in the *Journal of the American Medical Informatics Association* (JAMIA).

Citation

Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association* 2017;**24**(2):423-31 doi: 10.1093/jamia/ocw105

The version of record is available from the publisher's website:

https://doi.org/10.1093/jamia/ocw105.

2.4 Author contributions

David Lyell conceived this research and conducted the review, with guidance from, and under the supervision of, Enrico Coiera.

David Lyell drafted the manuscript, with revisions for intellectual content made by Enrico Coiera. Both authors approved the final manuscript.

Article I: Automation bias and verification complexity: a systematic review

Journal of the American Medical Informatics Association, 24(2), 2017, 423–431
doi: 10.1093/jamia/ocw105
Advance Access Publication Date: 11 August 2016
Review

Review

Automation bias and verification complexity: a systematic review

David Lyell and Enrico Coiera

Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Correspondence to David Lyell, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, NSW 2109, Australia; david.lyell@ students.mq.edu.au; Tel: +61 2 9850 2434

Received 9 March 2016; Revised 24 May 2016; Accepted 27 May 2016

ABSTRACT

Introduction: While potentially reducing decision errors, decision support systems can introduce new types of errors. Automation bias (AB) happens when users become overreliant on decision support, which reduces vigilance in information seeking and processing. Most research originates from the human factors literature, where the prevailing view is that AB occurs only in multitasking environments.
Objectives: This review seeks to compare the human factors and health care literature, focusing on the apparent association of AB with multitasking and task complexity.
Data sources: EMBASE, Medline, Compendex, Inspec, IEEE Xplore, Scopus, Web of Science, PsycINFO, and Business Source Premiere from 1983 to 2015.
Study selection: Evaluation studies where task execution was assisted by automation and resulted in errors were included. Participants needed to be able to verify automation correctness and perform the task manually.
Methods: Tasks were identified and grouped. Task and automation type and presence of multitasking were noted. Each task was rated for its verification complexity.
Results: Of 890 papers identified, 40 met the inclusion criteria; 6 were in health care. Contrary to the prevailing human factors view, AB was found in single tasks, typically involving diagnosis rather than monitoring, and with high verification complexity.
Limitations: The literature is fragmented, with large discrepancies in how AB is reported. Few studies reported the statistical significance of AB compared to a control condition.
Conclusion: AB appears to be associated with the degree of cognitive load experienced in decision tasks, and appears to not be uniquely associated with multitasking. Strategies to minimize AB might focus on cognitive load reduction.

Key words: decision support systems, clinical cognitive biases, complexity

INTRODUCTION

Automation in health care assists health professionals with complex or error-prone tasks such as diagnosis and treatment selection. For example, a clinical decision support system (CDSS) can help reduce prescribing errors by alerting clinicians to potential adverse events such as drug-drug interactions.[1] When it performs well, automation can reduce errors and improve decision performance.[1,2] It also, however, has the potential to introduce new types of errors.[3] One particularly significant risk is that users may become overreliant on automation, especially when a CDSS tool is less than perfectly accurate or reliable, leading to decision errors.[4]

This overreliance on less-than-perfect automation has been described in 2 separate but closely related bodies of research as either automation bias (AB) or automation-included complacency. Mosier and Skitka[5] define AB as "the tendency to use automated cues as a heuristic replacement for vigilant information seeking and

Box 1. Examples of automation bias in health care

Computer-aided detection (CAD) in radiology: CAD can help radiologists detect cancers in screening mammograms by placing prompts over suspicious image features. CAD has the potential to increase reader sensitivity and detect cancers that would otherwise be missed. However, there is also the risk that erroneous CAD prompts will result in cancers going undetected (omission errors), or patients without cancers being subjected to unnecessary and invasive testing or treatment (commission errors). This risk has been demonstrated in laboratory studies where qualified readers examined mammograms for the presence of cancers with and without the assistance of CAD. These studies found that when CAD failed to correctly prompt a cancer, subjects with CAD assistance performed significantly worse than unassisted readers. However, there was no significant difference in false positive prompts between the 2 groups. Hence, while radiologists were unaffected by false positive prompts from CAD, they were more likely to miss a cancer that was not prompted by CAD.[14,15]

Computerized EKG interpretation: Bogun et al.[16] found that, over a 6-month period in a major US hospital, 35% of patients with a diagnosis of atrial fibrillation were misdiagnosed by computerized interpretation of their EKGs. Of these, ordering physicians failed to correct the misdiagnosis for 24% of patients (a commission error), which led to an unnecessary change of treatment for 10% of patients, resulting in 2 patients (0.5%) developing complications.

processing." Here, automation provides cues for humans to attend to, and these are relied on more heavily than nonautomated cues. They distinguish between omission errors, where users fail to notice problems because they were not alerted to them by automation, and commission errors, where users act on incorrect advice given by automation. Automation-induced complacency[6] is "self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state."[7] Recently, Parasuraman and Manzey[8] reviewed both bodies of literature and argued that they are overlapping manifestations of the same automation-induced phenomenon, with allocation of limited user attention being central to both.

Despite the extensive use of automation and decision support in health care, little research has been explicitly conducted on AB. Some studies have documented AB associated with the use of CDSS without explicitly identifying the bias. Most existing research comes from human factors and ergonomics, where the prevailing view is that AB occurs only in environments where users perform multiple tasks simultaneously.[6,9]

However, studies from the health care literature, such as radiological computer-aided detection studies (Box 1), took place in single-task environments. A meta-analysis of AB in health care by Goddard et al.[10] found that incorrect decision support increased the risk of commission errors by 26% compared to when users did not have decision support. Three of the 4 studies were single task.[11–13]

Interestingly, high levels of system accuracy may inadvertently contribute to AB.[6,9,17–24] This may be because accuracy engenders trust, and it has been shown that users who have greater trust in automation are less likely to detect automation failures.[18,25]

Both task complexity[10,18] and task difficulty[15] have been cited as factors influencing AB, and are inherent properties of a task. To date, task complexity has been defined in terms of the cognitive demands of the task on the user[10,18] and the difficulty in terms of the portion of users who respond correctly to the task.[15] However, there is still no agreed on and systematic approach to the study of complexity in relation to AB, which limits a unified theoretical treatment of this phenomenon and also impedes the ability to design effective interventions to mitigate its effects.

This review seeks to compare and contrast the human factors and health care literature on automation bias, with a view to understanding the differences between the tasks in each and a specific focus on the apparent associations between AB, multitasking, and task complexity. The review will include studies of automation-induced complacency as well as AB, but for simplicity will use the term automation bias according to the definition provided by Mosier and Skitka.[5]

METHOD

A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-compliant systematic review was undertaken.[26] A literature search was conducted in July 2015 using EMBASE, Medline, Compendex, Inspec, IEEE Xplore, Scopus, Web of Science, PsycINFO, and Business Source Premiere.

The search consisted of medical subject headings (Appendix A) and the keywords "automation-induced complacency" and "automation bias." The search was limited to articles published in English since 1983. Eligible research studies needed to meet the following inclusion criteria:

- Experimental or observational research.
- Focus on the interaction between a human user and automation in performing a task.
- User had the capacity to perform the task manually without automation.
- User was presented with sufficient information to verify the correct functioning of automation.
- User had the ability to intervene in the task or choose when to use or rely on automation.
- Study tested the impact of an automation failure on human users in their performance of the task.

Study selection

Articles were screened using title and abstract, and those selected for full text assessment were then assessed by 2 reviewers (D.L. and R.S.). Interrater agreement was good (Cohen's κ 0.794; $n = 81$). Disagreements were resolved by consensus. Of 890 unique identified documents, 40 studies met all the inclusion criteria (Figure 1).

Quality assessment and risk of bias

Risk of bias was assessed using the Cochrane Collaboration tool for assessing risk of bias in randomized trials.[27] As the results of included studies were not pooled for meta-analysis, no papers were excluded on quality assessment.

A large portion of papers only provided summary details on recruitment, allocation, concealment, and blinding. Where data was missing or excluded, it was declared with reasons provided. Results were reported against hypotheses, but a sizable portion of studies did not report the results of statistical tests for significance of AB compared to a control condition.

Overall, most studies used relatively small samples (median 30, minimum 5, maximum 181). A large number of studies recruited

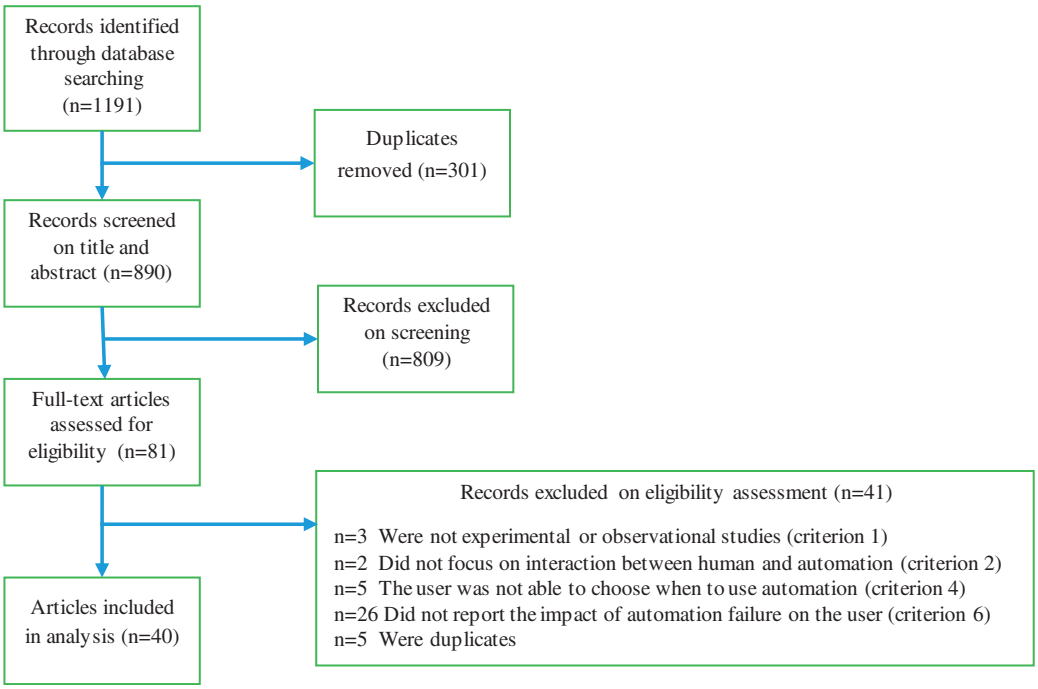

Figure 1. Selection of studies

university students, and nonstudent participants were typically recruited from the same or a small number of related organizations.

Data extraction and analysis

All experiments using the same task were grouped to allow for comparative analysis. Tasks were then systematically reviewed and data extracted for each of the following themes:

- Task: The experimental task being performed, including any secondary tasks.
- Task Type: The type of task being performed by the subject, classified as monitoring, diagnosis, and/or treatment.
  - Monitoring tasks require the user to monitor for a change in the state of a system. Usually this will be the transition from a desirable to a problematic state.
  - Diagnosis tasks require the user to identify and decide what the current state of the system is and/or what is causing the system to be in a problematic state.
  - Treatment tasks require the user to decide how to best treat or remedy the problem. Here the user attempts to change the system back to a desired state.
- Automation Type: Three categories of automated assistance were identified.
  - Alerting automation helps with monitoring tasks by alerting users to important changes in the state of a system.
  - Decision support assists users by providing a diagnosis of the problem or recommendations for treatment. At higher levels of automation, decision support may automatically implement recommendations.
  - Implementation automation assists users by implementing specified actions on their behalf; eg, air traffic control clearances sent via datalink can be automatically implemented directly into the flight management system. This category is used only when this occurs independent of decision support.
- Single-task or Multitask Environment: This refers to the number of different tasks that are performed simultaneously.

Verification complexity

For each study we sought to measure the complexity of using automation for a human. Specifically, we introduce the notion of verification complexity to describe the task complexity of verifying that automation is performing correctly. Verification actions can include assessing that alerts or recommendations are correct and ensuring that everything is satisfactory when no recommendations are being made by automation. For example, verifying a monitoring alert might require a user to observe that the alert is present only when a parameter is outside an acceptable range and never present when within normal operating range.

The verification complexity of each study was calculated by estimating the number of acquire, transform, interpret, or use steps necessary for a user to comprehensively test whether automation was functioning correctly (see Appendix B for an example). This was scored by 2 reviewers (D.L. and R.S.) and the intraclass correlation coefficient was 0.763, $F(36, 36) = 7.399$, $P < .001$, 95% CI (0.588-0.870), indicating a high degree of interrater reliability.

RESULTS

Forty studies met the inclusion criteria. These studies explored 17 different experimental tasks. The human factors literature produced 34 studies (85%) using 14 experimental tasks. The remaining 6 studies (15%) used 3 experimental tasks and came from the health care literature.

Automation bias

The majority of studies reported evidence of automation bias, with 81% of studies ($n = 25/31$) testing for omission errors and 91% ($n = 21/23$) testing for commission errors finding evidence of bias. Only 9 studies reported statistical significance when testing for the effect of AB against a nonautomated control (Table 1). Four of these found a significant effect for omission errors and 4 for commission errors. The ninth reported no significant effects for combined omission and commission errors. Effect sizes were not reported.

Some studies reported the significance of AB between different rates of automation accuracy. These consistently showed that participants made significantly more AB errors when assisted by automation that was constantly highly accurate compared to automation that varied between high and low accuracy.[6,19,20,24,25,28]

The remaining studies either reported the significance against interventions to mitigate AB or did not report on significance.

Task characteristics

The experimental tasks could be divided into 17 unique tasks. The 3 tasks from the health care literature were computer-aided detection of cancers in screening mammography, computerized interpretation of EKGs, and computerized clinical decision support systems. The remaining 14 experimental tasks came from the human factors literature. Nine tasks originated from aviation, and 1 task each came from process control, military, security, nuclear power, and space.

Eleven of the tasks, all originating from the human factors literature, required subjects to perform 2, 3, or 4 tasks concurrently, with 3 being the most common number of concurrent tasks. The remaining 6 experimental tasks involved a single task, 3 of which originated from the health care literature and 3 from the human factors literature. All experimental tasks reported in the health care literature were single task. Interestingly, there were 2 studies that compared the same subtask across single-task and multitask conditions, which found evidence of omission errors in the multitask but not in the single-task condition.[6,9] Single-task studies involved mostly diagnosis, whereas multitask studies included all 3 task types.

Verification complexity

In multitasks experiments, all subtasks assisted by automation were assessed for verification complexity (Table 2). Ratings varied between low and high: 4 low, 2 medium, and 5 high. Tasks across all levels produced evidence of automation bias. Similarly, single-task experiments also varied between low and high: 1 low, 2 medium, and 4 high. However, only tasks rated medium or higher produced automation bias. Two studies describing single tasks did not contain sufficient information to allow for an assessment of demands on working memory.[31,52]

DISCUSSION

Single-task vs multitask

The prevailing view in the human factors literature is that automation bias occurs only in multitask rather than single-task environments.[6] Consistent with this, 10 of the 11 multitask studies reported automation bias. However, we also found 5 of 6 single-task studies that produced evidence of automation bias. Two of these, a luggage screening task[51] and the nondestructive testing of components in nuclear power plants,[52] came from the human factors literature. All 3 of the experimental tasks from health care were single task.

This finding signals a key point of difference between the human factors and health care literature. It also has substantial theoretical implications. First, it represents a significant departure from the prevailing view that automation bias occurs only in multitask environments. Second, the finding challenges the theoretical proposition put forward by Parasuraman and Manzey[8] that automation bias occurs when multiple tasks compete for the user's attention.

Task and automation type

The primary task for most multitask experiments required subjects to monitor for changes in a system or decide how to best manage the problem. Monitoring tasks were assisted by alerts that notified users of a change in system state. Treatment tasks were assisted by decision support, which provided recommendations for remedies. Monitoring requires a user to detect a change from a desirable to an undesirable state. Usually this involves monitoring changes in parameters over time. Examples include monitoring engine gauges to determine whether there was a change from being within tolerances to exceeding them.

In contrast, all but 1 single-task experiment involved diagnosis, which requires the subject to ascertain the current state of the system, assisted by decision support. Unlike monitoring, diagnosis was not concerned with detecting a change in the system over time. Examples include viewing mammograms and determining whether or not a cancer is present.

Monitoring tasks were used in studies to investigate omission errors, while diagnosis and treatment tasks were used to investigate commission errors.

The 2 human factors studies that did report evidence of automation bias in a single-task environment both involved diagnosis. The first was a luggage screening task,[51] which required subjects to view an X-ray image and then decide whether or not a weapon was present. The second, the nondestructive testing of components in a nuclear power plant,[52] required subjects to check computerized interpretations of eddy current testing on components.

Verification complexity

Our analysis showed that single-task experiments that reported evidence of automation bias were rated medium to high for verification complexity. The picture for multitasking studies was similar, in that multiple low-complexity tasks could combine to generate automation bias. Two Multi-Attribute Task Battery experiments with low verification complexity did not produce automation bias until they were combined with a secondary task.[6,9]

This suggests that (1) a higher level of verification complexity is required for automation bias to present in a single-task than a multitask setting, and (2) the cognitive demands of tasks are cumulative; ie, the addition of secondary tasks appears to increase demands on a user to the point where errors emerge.

Very little research has been conducted on the relationship between task complexity and automation bias. No studies have directly compared task complexity between single-task and multitask settings. However, 2 studies reported that high task complexity or task difficulty resulted in more automation bias errors, providing some support for our observation that differences in task complexity may explain why some single tasks produce automation bias while others do not. For example, Bailey and Scerbo[18] manipulated the complexity of a monitoring task, finding that subjects made more automation bias errors when performing more complex (ie, more cognitively demanding) monitoring tasks.

The role of task complexity is also partially supported by Wickens and Dixon,[54] who, in a review of the costs and benefits of imperfect automation, found that user dependence on automation was greater under conditions of high workload, which they defined in terms of task difficulty or concurrent task load. They confirmed this finding in a laboratory experiment that suggests dependence on imperfect automation is greater under circumstances of high workload when user resources are assumed to be scarce.[55] However, this poses additional risk, since once errors are made, they are less likely to be detected under conditions of high workload.[56]

Table 1. Characteristics of experimental tasks and the reported significance of automation bias

Task	Single or Multi	Subtasks	Task Type	Automation Type	Study	Sample	Trials	Omission Errors	Commission Errors
Mammography, computer-aided detection	Single	Screen mammograms for cancers	D	DS	(14)	19 readers	60 sets of mammograms	$P < .000001^{\#}$	–
					(29)	5 readers	185 sets of mammograms	–	Not reported
					(15)	44 readers	180 mammograms	–	Not reported
EKG, computerized interpretation	Single	Diagnosis of atrial fibrillation	D	DS	(16)	2298 EKGs from 1085 patients		–	Not reported
Clinical decision support system	Single	Prescribe treatment for patient scenarios	T	DS	(30)	26 general practitioners	20 scenarios	–	$P < .05^{\#}$
Clinical decision support system	Single	Answer clinical questions	D	DS	(31)	29 general practitioners	15 questions about clinical cases	–	$P = .031^{\#}$
Multi-Attribute Task Battery (multitask)	Multi	System monitoring task	M	A	(17)	24 participants	12 * 10 mins	$P > .05^{\#\#\#}$	–
					(25)	24 engineering students	12 * 10 mins	$P < .001^{\#\#}$	–
					(19)	40 undergraduate students	3 * 15 mins	$P = .013^{\#\#}$	–
					(32)	16 students	4 * 15 mins	–	Not reported
		Tracking task[a]			(9)	24 students	4 * 30 mins	$P < .05^{\#\#}$	–
					(6)	24 participants	12 * 10 mins	$P < .0001^{\#\#}$	–
					(20)	40 undergraduate students	2 * 40 mins	$P < .0001^{\#\#}$	–
					(22)	120 students	6 * 10 mins	Not reported	–
		Fuel management task[a]			(33)	20 students	6 * 10 mins	$P > .05^{\#\#\#}$	–
					(23)	120 subjects	6 * 10 mins	$P < .01^{\#\#}$	–
					(34)	16 students	3 * 10 mins	$P < .01^{\#}$	–
					(24)	80 participants	6 * 10 mins	$P < .01^{\#\#}$	–
Multi-Attribute Task Battery (single task)	Single	System monitoring task	M	A	(9)	16 students	4 * 30 mins	$P > .05^{\#\#}$	–
					(6)	16 adults	12 * 10 mins	$P > .05^{\#\#}$	–
Workload/PerformANcE Simulation (W/PANES)	Multi	Gauge monitoring task	M, T	A, DS	(35)	181 students	5 * 10 mins	$P < .001^{\#\#\#}$	$P < .001^{\#\#\#}$
		Tracking task[a]			(36)	80 students	8 * 8 mins	$P < .05^{\#}$	Not reported
		Waypoints task[a]			(37)	144 students	4 trials	$P > .05^{\#\#\#}$	$P < .05^{\#\#\#}$
mini-Advanced Concepts Flight Simulator (ACFS)	Multi	Datalink clearances	M	I	(38)	48 commercial pilots	3 part flights	Not reported	Not reported
		EICAS event (engine fire)	M, D, T	A, DS	(39)	25 commercial pilots	2 part flights	Not reported	Not reported
		Tracking task[a]							
Smart icing system	Multi	Recover from inflight icing events	D, T	DS	(40)	30 flight instructors	28 part flights	–	Not reported
		Respond to air traffic control[a]			(41)	27 commercial pilots	10 part flights	–	$P < .001^{\#}$
		Monitor instruments for failures[a]							
NASA Stone Soup simulator	Multi[b]	Datalink clearances	M	I	(42)	30 commercial pilots	8 * 10 min part flights	–	Not reported
Airbus A320 touch screen trainer	Multi[b]	Detect throttle malfunction	M	A	(43)	35 pilots	1 * 40 mins	Not reported	–
Unnamed part flight task	Multi	Gauge monitoring task	M	A	(18)	32 participants	1 * 100 mins	$P \leq .05^{\#\#}$	–
		Mode monitoring task	M	A					

(continued)

Table 1. Continued

Task	Single or Multi	Subtasks	Task Type	Automation Type	Study	Sample	Trials	Omission Errors	Commission Errors
		Digital display monitoring task	M						
		Tracking task[a]							
Air traffic control task	Multi	Aircraft conflict detection	M	A	(44)	20 air traffic controllers	5 * 25 mins	$P < .05^{\#}$	–
		Communications task[a]							
		Updating flight strips[a]							
Pilot air traffic conflict detection	Single	Estimate the closest point of approach between own and intruder aircraft	D	A	(45)	24 pilots	72 trials	$P = .46^{\#}$	–
AutoCAMS	Multi	Supervisory control of life support system (eg, spacecraft)	D, T	DS	(46)	24 engineering students	1 * 75 mins	–	Not reported
					(47)	24 engineering students	1 * 100 mins	$P < .01^{\#\#\#}$	Not reported
		Prospective memory task[a]			(48)	88 engineering students	5 * 40 mins	–	$P < .03^{\#\#\#}$
		Communication task[a]			(49)	88 engineering students	5 * 40 mins	–	$P < .03^{\#\#\#}$
					(50)	32 engineering students	5 * 40 mins	–	$P < .02^{\#\#\#}$
Command and control: sensor to shooter	Multi	Making friendly-enemy engagement decisions	D, T	DS	(28)	100 adults	300 scenarios	–	$P < .0001^{\#\#\#}$
		Communications task[a]			(21)	18 undergraduate students	500 scenarios	–	$P < .001^{\#}$
Baggage X-ray screening	Single	Detection of weapons in X-rays of luggage	D	DS	(51)	96 undergraduate students	200 images	$P < .001^{\#\#}$	Not reported
Nondestructive testing in nuclear power plants	Single	Detection and sizing of faults	D	DS	(52)	70 trainees	36 trials	–	Not reported
Robotic arm simulator	Multi	Control of robotic arm to target location	M, T	A, DS	(53)	36 participants	7 trials	$P < .01^{\#\#\#}$	$P < .01^{\#\#\#}$
		Camera selection	T	DS					

Task type: M = monitoring; D = diagnosis; T = treatment.
Automation type: A = alerting; I = implementation; DS = decision support.
[a]Secondary tasks not assessed for automation bias. [b]The task was performed in a multitask environment, but the article did not specify the secondary tasks.
[#]Probability of automation bias errors compared to manual control.
[##]Probability of automation bias errors between different levels of automation accuracy.
[###]Probability of automation bias errors for tested intervention.
Not reported = Probability for automation bias errors not reported.
– = Not tested.

Cognitive load theory: a framework for addressing task complexity

Cognitive load theory may help to explain these findings.[57] This theory has been developed by educational psychologists over the last 30 years with the aim of improving learning outcomes. Central to the theory is the notion of limited capacity of human working memory. It claims that the learning process requires students to manipulate information in working memory, generating a cognitive load. Learning fails when the cognitive load generated by the task exceeds the student's available working memory. Cognitive load theorists differentiate intrinsic and extraneous cognitive load.[57] Intrinsic cognitive load is generated by the task being learned, whereas extraneous cognitive load arises from other sources unrelated to the learning task.

One of the stated aims of cognitive load theory is to develop interventions that reduce extraneous cognitive load and allow more resources to be allocated to learning. This theory may be applicable to human-automation interactions in order to understand the cognitive demands that work with automation places on

Table 2. Verification complexity

Task	Single or Multi	Subtasks	Verification complexity
Mammography, computer-aided detection[14,15,29]	Single	Screen mammograms for cancers	High[#]
EKG, computerized interpretation [16]	Single	Diagnosis of atrial fibrillation	High
Clinical decision support system[30]	Single	Prescribe treatment for patient scenarios	High[#]
Clinical decision support system[31]	Single	Answer clinical questions[c]	
Multi-Attribute Task Battery (multitask)[6,9,17,19,20,22–25,32–34]	Multi	System monitoring task	Low
		Tracking task[a]	Low
		Fuel management task[a]	Medium
Multi-Attribute Task Battery (single task)[6,9]	Single	System monitoring task	Low
Workload/PerformANcE Simulation (W/PANES)[35–37]	Multi	Gauge monitoring task	Low
		Tracking task[a]	Low
		Waypoints task[a]	Low
mini-Advanced Concepts Flight Simulator (ACFS)[38,39]	Multi	Datalink clearances	Low
		EICAS event (engine fire)	Low
		Tracking task[a]	Low
Smart icing system[40,41]	Multi	Recover from inflight icing events	High
		Respond to air traffic control[a]	Low
		Monitor instruments for failures[a]	Medium
NASA Stone Soup simulator [42]	Multi[b]	Datalink clearances	Medium
Airbus A320 touch screen trainer[43]	Multi[b]	Detect throttle malfunction	Low
Unnamed part flight task[18]	Multi	Gauge monitoring task	Low
		Mode monitoring task	Medium
		Digital display monitoring task	High
		Tracking task[a]	Low
Air traffic control task[44]	Multi	Aircraft conflict detection	High
		Communications task[a]	Medium
		Updating flight strips[a]	High
Pilot air traffic conflict cetection[45]	Single	Estimate the closest point of approach between own and intruder aircraft	Medium
AutoCAMS[46–50]	Multi	Supervisory control of life support system (eg, spacecraft)	High
		Prospective memory task[a]	Low
		Communication task[a]	Low
Command and control: sensor to shooter[21,28]	Multi	Making friendly-enemy engagement decisions	Medium
		Communications task[a]	Low
Baggage X-ray screening[51]	Single	Detection of weapons in X-rays of luggage	Medium[#]
Nondestructive testing in nuclear power plants[52]	Single	Detection and sizing of faults[c]	
Robotic arm simulator[53]	Multi	Control of robotic arm to target location	High
		Camera selection	Medium

[a]Secondary tasks not assessed for automation bias.
[b]The task was performed in a multitask environment, but the article did not specify the secondary tasks.
[c]The article did not present sufficient information to assess the verification complexity of the task.
[#]Tasks containing a variable or unspecified number of information cues. An estimation was used in the analysis.

users. We hypothesize that, just as cognitive overload can prevent learning, it can also prevent users from being able to adequately verify the correct operation of automation and lead to automation bias errors.

The cognitive load theory framework can also apply to the verification of automation in the same way. It is possible that supervision of automation generates both an intrinsic and extraneous cognitive load. Intrinsic load would come from the cognitive demands of verifying automation, and extraneous load might come from the manner in which information is presented or from the sociotechnical environment in which the primary task is being conducted.

More broadly, cognitive overload may also explain the observed discrepancy between the findings of human factors and health care researchers. Whereas the former found automation bias only in multitask settings, the latter found evidence of automation bias in single-task settings, but these tasks were likely to be associated with a higher cognitive load.

Implications

Decision-making in health care is complex, as it is characterized by high levels of ambiguity and detail (eg, in mammograms) and/or large volumes of information (eg, in drug-drug interactions). Computerized decision support systems thus play a central role in this setting by helping health care professionals manage this complexity.[58] CDSS alerts can provide an opportunity to detect and recover from errors that have been missed by clinicians. Likewise, verification is a vital step that allows for the detection of and recovery from CDSS failures. However, a major obstacle to this is the complexity of the task of verification. High verification complexity appears to increase the risk of automation bias by increasing cognitive load, making it difficult for health care professionals to verify CDSS performance.

To date, interventions designed to counter automation bias have had little or no impact. Interventions tested thus far have manipu-

lated user accountability for performance, which had only a mild effect on novice subjects and no effect on expert subjects.[35] Providing subjects with contextual information on the reliability of automation reduced automation bias in some, but not all, experimental conditions.[17] Providing subjects with feedback on performance had no impact,[3] and training interventions resulted in no significant reduction in rates of automation bias. These included providing additional training in performing the task manually,[22] exposing users to examples of automation failure,[46] and providing explicit training on automation bias and how to avoid errors.[38]

If automation bias is partly due to cognitive demands that exceed the user's capacity, interventions seeking to reduce automation bias that do not address this cognitive overload are unlikely to succeed. Indeed, the interventions reported in the included studies have produced, at best, marginal reductions in the rates of automation bias.

Efforts can focus on reducing cognitive load from sources external to the task in busy clinical environments. For example, eliminating distractions by introducing no-interruption zones[59] or improving the fit between decision support with existing workflows using simplified user interfaces[60] should free up additional cognitive resources to attend to the primary task. It will also be necessary to target the cognitive load generated by the task of verification itself. Designing interfaces that support effective verification, eg, by presenting critical verification information side-by-side with decision support, may help in this respect.

Limitations

There is significant fragmentation of the automation bias literature, not just in terms of the divide between automation bias and automation-induced complacency, but also because many studies reporting automation bias do not identify with either camp.[29]

Comparing studies was difficult.[10] Only 9 studies reported the significance for the presence of automation bias compared to a manual (nonautomated) control. This, combined with the large variability in the reported measures, makes it difficult to draw comparisons between studies. Studies used very homogeneous samples of subjects, few in number, and mostly university students or professionals recruited from the same or a small number of closely related organizations.

CONCLUSION

This review set out to compare and contrast different types of experimental tasks reported in the human factors and health care literature. We found, contrary to the prevailing view within the human factors literature, which holds that automation bias occurs only in a multitask environment, that all the health care experiments reporting automation bias took place in a single-task environment.

To understand this contradiction and why some single tasks produced automation bias while others did not, we examined the characteristics of the experimental tasks. Our analysis revealed that single tasks that produced automation bias had higher verification complexity than single tasks that did not.

Cognitive load theory provides a robust framework for studying the impact of task/verification complexity on automation bias. Further research is needed to test these hypotheses, especially in terms of the impact of cognitive load on automation bias and the potential of cognitive load theory to explain why human factors experiments showed no evidence of automation bias in a single-task setting, whereas health care experiments did.

ACKNOWLEDGMENTS

Rhonda Siu assisted in the screening of articles for inclusion in the review and the scoring of verification complexity in experimental tasks.

CONTRIBUTORS

D.L. conceived this research and conducted the review, with guidance from and under the supervision of E.C.

D.L. drafted the manuscript, with revisions for intellectual content made by E.C. Both authors approved the final manuscript.

FUNDING

This research was supported by a doctoral scholarship for D.L. provided by the HCF Research Foundation.

COMPETING INTERESTS

The authors have no competing interests to declare.

SUPPLEMENTARY MATERIAL

Supplementary material are available at *Journal of the American Medical Informatics Association* online.

REFERENCES

1. Wolfstadt JI, Gurwitz JH, Field TS, *et al*. The effect of computerized physician order entry with clinical decision support on the rates of adverse drug events: a systematic review. *J General Int Med* 2008;23(4):451–458.
2. Garg AX, Adhikari NJ, McDonald H, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;293(10):1223–1238.
3. Coiera E, Westbrook J, Wyatt J. The safety and quality of decision support systems. *Methods Inf Med* 2006;45(Suppl 1):20–25.
4. Coiera E. Technology, cognition and error. *BMJ Qual Saf* 2015;24(7):417–422.
5. Mosier KL, Skitka LJ. Human decision makers and automated decision aids: made for each other. In: R Parasuraman, M Mouloua, eds. *Automation and Human Performance: Theory and Applications*. Hillsdale, NJ, England: Lawrence Erlbaum Associates; 1996:201–220.
6. Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced "complacency." *Int J Aviation Psychol. 1993*;3(1):1–23.
7. Billings C, Lauber J, Funkhouser H, Lyman G, Huff E. *NASA Aviation Safety Reporting System*. (Technical Report TM-X-3445). Moffett Field, Calif.: NASA Ames Research Center; 1976.
8. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. *Human Factors* 2010;52(3):381–410.
9. Molloy R, Parasuraman R. Monitoring automation failures: effects of automation reliability and task complexity. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. October 12-16, 1992. Atlanta, Ga.: Human Factors and Ergonomics Society; 1992:1518–1521.
10. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc* 2012;19(1):121–127.
11. Hillson SD, Connelly DP, Yuanli Liu. The effects of computer-assisted electrocardiographic interpretation on physicians' diagnostic decisions. *Med Decis Mak* 1995;15(2):107–112.
12. Southern WN, Arnsten JH. The Effect of Erroneous Computer Interpretation of ECGs on Resident Decision Making. *Med Decis Mak* 2009;29(3):372–376.
13. Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. *J Am Med Inform Assoc* 2003;10(5):478–483.

14. Alberdi E, Povykalo A, Strigini L, Ayton P. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. *Acad Radiol* 2004;11(8):909–918.
15. Povyakalo AA, Alberdi E, Strigini L, Ayton P. How to discriminate between computer-aided and computer-hindered decisions: a case study in mammography. *Med Decis Mak* 2013;33(1):98–107.
16. Bogun F, Anh D, Kalahasty G, *et al*. Misdiagnosis of atrial fibrillation and its clinical consequences. *Am J Med* 2004;117(9):636–642.
17. Bagheri N, Jamieson GA. The impact of context-related reliability on automation failure detection and scanning behaviour. *Conf Proc IEEE Int Conf Syst Man Cybern*, Vol. 1. October 10-13, 2004. 2004:212–217.
18. Bailey NR, Scerbo MW. Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues Ergonomics Sci* 2007;8(4):321–348.
19. Bailey NR, Scerbo MW, Freeman FG, Mikulka PJ, Scott LA. Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors* 2006;48(4):693–709.
20. Prinzel LJ III, Freeman FG, Prinzel HD. Individual differences in complacency and monitoring for automation failures. *Individual Differences Res* 2005;3(1):27–49.
21. Rovira E, McGarry K, Parasuraman R. Effects of imperfect automation on decision making in a simulated command and control task. *Hum Factors* 2007;49(1):76–87.
22. Singh AL, Tiwari T, Singh IL. Effects of automation reliability and training on automation-induced complacency and perceived mental workload. *J Indian Acad Appl Psychol* 2009;35(Spec iss):9–22.
23. Singh IL, Sharma HO, Parasuraman R. Effects of Training and Automation Reliability on Monitoring Performance in a Flight Simulation Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 2000;44(13):53–56.
24. Singh IL, Singh AL, Saha PK. Monitoring performance and mental workload in an automated system. *Proceedings of the International Conference on Engineering Psychology and Cognitive Ergonomics*, July 22-27, 2007. Beijing, China: Springer Verlag; 2007: 426–435.
25. Bagheri N, Jamieson GA. *Considering Subjective Trust and Monitoring Behavior in Assessing Automation-induced "Complacency."* Mahwah, N.J.: Lawrence Erlbaum Assoc; 2004.
26. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann Int Med* 2009;151(4):264–269.
27. Higgins JPT, Altman DG, Gøtzsche PC, *et al*. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:55:48.
28. Parasuraman R, de Visser E, Lin M-K, Greenwood PM. Dopamine beta hydroxylase genotype identifies individuals less susceptible to bias in computer-assisted decision making. *PLoS ONE* 2012;7(6):e39675.
29. Marx C, Malich A, Facius M, *et al*. Are unnecessary follow-up procedures induced by computer-aided diagnosis (CAD) in mammography? Comparison of mammographic diagnosis with and without use of CAD. *Eur J Radiol* 2004;51(1):66–72.
30. Goddard K, Roudsari A, Wyatt JC. Automation bias: empirical results assessing influencing factors. *Int J Med Inform* 2014;83(5):368–375.
31. Golchin K, Roudsari A. Study of the effects of clinical decision support system's incorrect advice and clinical case difficulty on users' decision making accuracy. *Stud Health Technol Inform* 2011;164:13–16.
32. Harris WC, Goernert PN. The effect of levels of automation on supervisory performance in a multi-task environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, October 5-9, 1998. Santa Monica, Calif.: Human Factors and Ergonomics Society; 1998: 128–132.
33. Singh AL, Tiwari T, Singh IL. Performance feedback, mental workload and monitoring efficiency. *J Indian Acad Appl Psychol* 2010;36(1):151–158.
34. Singh IL, Sharma HO, Parasuraman R. Effects of manual training and automation reliability on automation induced complacency in flight simulation task. *Psychol Stud* 2001;46(1-2):21–27.
35. Skitka LJ, Mosier K, Burdick MD. Accountability and automation bias. International *J Hum Comp Stud* 2000;52(4):701–717.
36. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making? *Int J Hum Comp Stud* 1999;51(5):991–1006.
37. Skitka LJ, Mosier KL, Burdick M, Rosenblatt B. Automation bias and errors: are crews better than individuals? *Int J Aviat Psychol* 2000;10(1):85–97.
38. Mosier KL, Skitka LJ, Dunbar M, McDonnell L. Aircrews and automation bias: The advantages of teamwork? *Int J Aviat Psychol* 2001;11(1):1–14.
39. Mosier KL, Skitka LJ, Heers S, Burdick M. Automation bias: decision making and performance in high-tech cockpits. *Int J Aviat Psychol* 1998;8(1):47–63.
40. McGuirl JM, Sarter NB. Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Hum Factors* 2006;48(4):656–665.
41. Sarter NB, Schroeder B. Supporting decision making and action selection under time pressure and uncertainty: the case of in-flight icing. *Hum Factors* 2001;43(4):573–583.
42. Olson WA, Sarter NB. Management by consent in human-machine systems: when and why it breaks down. *Human Factors* 2001;43(2):255–266.
43. de Boer RJ, Heems W, Hurts K. The duration of automation bias in a realistic setting. *Int J Aviat Psychol* 2014;24(4):287–299.
44. Metzger U, Parasuraman R. Automation in future air traffic management: effects of decision aid reliability on controller performance and mental workload. *Human Factors* 2005;47(1):35–49.
45. Xu X, Wickens CD, Rantanen EM. Effects of conflict alerting system reliability and task difficulty on pilots' conflict detection with cockpit display of traffic information. *Ergonomics* 2007;50(1):112–130.
46. Bahner J, Huper A-D, Manzey D. Misuse of automated decision aids: complacency, automation bias and the impact of training experience. *Int J Hum Comp Stud* 2008;66(9):688–699.
47. Bahner JE, Elepfandt MF, Manzey D. Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, September 22-26, 2008. New York: Human Factors and Ergonomics Society; 2008: 1330–1334.
48. Manzey D, Reichenbach J, Onnasch L. Human performance consequences of automated decision aids: the impact of degree of automation and system experience. *J Cogn Eng Decis Mak* 2012;6(1):57–87.
49. Reichenbach J, Onnasch L, Manzey D. Misuse of automation: the impact of system experience on complacency and automation bias in interaction with automated aids. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, September 27 to October 1, 2010. San Francisco, Calif.: Human Factors and Ergonomics Society; 2010: 374–378.
50. Reichenbach J, Onnasch L, Manzey D. Human performance consequences of automated decision aids in states of sleep loss. *Human Factors* 2011;53(6):717–728.
51. Rice S, McCarley JS. Effects of response bias and judgment framing on operator use of an automated aid in a target detection task. *J Exp Psychol* 2011;17(4):320–331.
52. Bertovic M, Fahlbruch B, Mueller C. Human factors perspective on the reliability of NDT in nuclear applications. *Mater Test* 2013;55(4):243–253.
53. Huiyang L, Wickens CD, Sarter N, Sebok A. Stages and levels of automation in support of space teleoperations. *Hum Factors* 2014;56(6):1050–1061.
54. Wickens CD, Dixon SR. The benefits of imperfect diagnostic automation: a synthesis of the literature. *Theoretical Issues Ergonomics Sci* 2007;8(3):201–212.
55. Dixon SR, Wickens CD. Automation reliability in unmanned aerial vehicle control: a reliance-compliance model of automation dependence in high workload. *Hum Factors* 2006;48(3):474–486.
56. Amalberti R, Wioland L. Human error in aviation. In: Soekkha HM, editor. Aviation Safety. The Netherlands: VSP; 1997. p. 91-108.
57. Sweller J, Ayres P, Kalyuga S. *Cognitive Load Theory*. New York: Springer; 2011.
58. Sintchenko V, Coiera EW. Which clinical decisions benefit from automation? A task complexity approach. *Int J Med Inform* 2003;70(2–3):309–316.
59. Coiera E. The science of interruption. *BMJ Qual Saf* 2012;21(5):357–360.
60. Goddard K. *Automation Bias and Prescribing Decision Support–rates, Mediators and Mitigators*. Unpublished doctoral thesis, City University London; 2012.

Article I: Appendices

APPENDIX A: MEDICAL SEARCH HEADINGS SEARCH TERMS

Medical Subject Heading (MeSH) terms used in the literature search are shown in Table 1. Searches comprised three components: (1) a task, (2) the use of automation to assist in the task, and (3) errors occurring in the task as a result of automation.

<table>
<tr><th>Task</th><th>Automation</th><th>Error</th></tr>
<tr><td>decision making
diagnosis
diagnosis, differential
diagnostic techniques and procedures
incidental findings
prognosis
task performance analysis</td><td>automation
attitude to computer
user-computer interface
man-machine systems</td><td rowspan="3">medical errors
diagnostic errors
medication errors
inappropriate prescribing</td></tr>
<tr><th colspan="2">Task and Automation</th></tr>
<tr><td colspan="2">diagnosis, computer-assisted
image processing, computer-assisted
surgery, computer-assisted
decision making, computer-assisted
diagnosis, computer-assisted
therapy, computer-assisted
decision support systems, clinical</td></tr>
</table>

Table 1 Medical Subject Heading search terms

APPENDIX B: ESTIMATING VERIFICATION COMPLEXITY

The estimated number of steps the user needs to perform in order to complete the task. Each step represents one action that is taken in order to *acquire*, *transform*, *interpret* and *use* information. This assumes that the more steps are involved, and the more information the subject needs to attend to, the more complex the task. Each step, no matter what type of action, adds a score of one to the task complexity, in line with similar complexity scores.

For example, if a subject performing the unnamed part flight simulator task, has to determine if the EGT (engine exhaust temperature) value on a digital display is within tolerances, this would involve four steps. If this task were assisted by automation, verifying it would require the subject to compare the result they obtained with that of decision support, an additional two steps.

Step 1. **Acquire:** Remember the lower threshold of 320.
Step 2. **Acquire:** Remember the upper threshold of 340.
Step 3. **Acquire:** Acquire the EGT value.
Step 4. **Interpret:** Compare threshold values for EGT to actual EGT.
Step 5. **Acquire:** Acquire decision support outcome.
Step 6. **Interpret:** Compare the actual decision support outcome with expected outcome.

The user's level of expertise moderates how many steps are involved. Through learning an expert constructs schema stored in long term memory which they can draw upon. For an experienced pilot performing the same task might only require four steps.

Step 1. **Acquire:** Acquire EGT value
Step 2. **Interpret:** Compare EGT value with schema for the acceptable EGT range.
Step 3. **Acquire:** Acquire decision support outcome.
Step 4. **Interpret:** Compare the actual decision support outcome with expected outcome.

In this analysis, if the study participants where described as qualified in the task they were performing (for example qualified radiologists or pilots), they were taken to be experts. Otherwise they were treated as novices.

The number of steps were grouped into low, medium and high according to the following.

Low	Less than 9 steps.
Medium	Between 10 and 19 steps.
High	Greater than 20 steps.

2.5 Chapter 2 References

1. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 2012;**19**(1):121-27 doi: 10.1136/amiajnl-2011-000089

2. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association* 2017;**24**(2):423-31 doi: 10.1093/jamia/ocw105

3. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: An attentional integration. *Human Factors* 2010;**52**(3):381-410 doi: 10.1177/0018720810376055

4. Bagheri N, Jamieson GA. The impact of context-related reliability on automation failure detection and scanning behaviour. IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583); 2004 Oct 10-13.

5. Bagheri N, Jamieson GA. Considering subjective trust and monitoring behavior in assessing automation-induced "complacency". In: Vincenzi DA, Mouloua M, Hancock PA, eds. Human Performance, Situation Awareness and Automation: Current Research and Trends, Vol 2. Mahwah: Lawrence Erlbaum Associates, 2004:54-59.

6. Bailey NR, Scerbo MW, Freeman FG, Mikulka PJ, Scott LA. Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors* 2006;**48**(4):693-709 doi: 10.1518/001872006779166280

7. Molloy R, Parasuraman R. Monitoring automation failures: effects of automation reliability and task complexity. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 1992 Oct 12-16; Atlanta, GA, USA. Human Factors And Ergonomics Society.

8. Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced "complacency.". *The International Journal of Aviation Psychology* 1993;**3**(1):1-23 doi: 10.1207/s15327108ijap0301_1

9. Prinzel LJ, III, Freeman FG, Prinzel HD. Individual Differences in Complacency and Monitoring for Automation Failures. *Individual Differences Research* 2005;**3**(1):27-49

10. Singh AL, Tiwari T, Singh IL. Effects of automation reliability and training on automation-induced complacency and perceived mental workload. *Journal of the Indian Academy of Applied Psychology* 2009;**35**(spec iss):9-22

11. Singh AL, Tiwari T, Singh IL. Performance feedback, mental workload and monitoring efficiency. *Journal of the Indian Academy of Applied Psychology* 2010;**36**(1):151-58

12. Singh IL, Sharma HO, Parasuraman R. Effects of Training and Automation Reliability on Monitoring Performance in a Flight Simulation Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 2000;**44**(13):53-56 doi: 10.1177/154193120004401314

13. Singh IL, Sharma HO, Parasuraman R. Effects of manual training and automation reliability on automation induced complacency in flight simulation task. *Psychological Studies* 2001;**46**(1/2):21-27

14. Singh IL, Singh AL, Saha PK. Monitoring performance and mental workload in an automated system. Proceedings of the International Conference onEngineering Psychology and Cognitive Ergonomics; 2007 Jul 22-27; Beijing, China. Springer Verlag.

15. Skitka LJ, Mosier K, Burdick MD. Accountability and automation bias. *International Journal of Human-Computer Studies* 2000;**52**(4):701-17 doi: 10.1006/ijhc.1999.0349

16. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making? *International Journal of Human Computer Studies* 1999;**51**(5):991-1006 doi: 10.1006/ijhc.1999.0252

17. Skitka LJ, Mosier KL, Burdick M, Rosenblatt B. Automation bias and errors: Are crews better than individuals? *The International Journal of Aviation Psychology* 2000;**10**(1):85-97 doi: 10.1207/S15327108IJAP1001_5

18. Bahner J, Huper A-D, Manzey D. Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies* 2008;**66**(9):688-99 doi: 10.1016/j.ijhcs.2008.06.001

19. Bahner J, Elepfandt MF, Manzey D. Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2008 Sep 22-26; New York, NY, United states. Human Factors And Ergonomics Society.

20. Manzey D, Reichenbach J, Onnasch L. Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience. *Journal of Cognitive Engineering and Decision Making* 2012;**6**(1):57-87 doi: 10.1177/1555343411433844

21. Reichenbach J, Onnasch L, Manzey D. Misuse of automation: The impact of system experience on complacency and automation bias in interaction with automated aids. Proceedings of the Human

Factors and Ergonomics Society Annual Meeting; 2010 Sep 27 - Oct 1; San Francisco, CA, United states. Human Factors And Ergonomics Society.

22. Reichenbach J, Onnasch L, Manzey D. Human performance consequences of automated decision aids in states of sleep loss. *Human Factors* 2011;**53**(6):717-28 doi: 10.1177/0018720811418222

3 Automation bias and error

This chapter focuses on the relationship between automation bias and errors.

3.1 Background

The healthcare studies included in the systematic review (chapter 2) involved decision-making tasks which were assisted by computerised decision support and took place within a single task environment.[1] The review suggested that increasing task complexity may be a key risk factor for automation bias in healthcare tasks.[1] An experiment was conducted to: (1) test for the presence of automation bias in a clinical decision-making task assisted by computerised clinical decision support, and (2) determine whether increased task complexity is a risk factor for automation bias. These will be assessed by examining the effect of task complexity and clinical decision support on the key outcomes of errors, cognitive load (information processing) and verification (information seeking).

The effect of automation bias was established by comparing an experimental condition that provided opportunities for participants to make automation bias errors and a control condition involving no decision support. If participants suffered an automation bias, there would have been more errors in the experimental than in the control conditions.

Prescribing errors inserted into the experimental task provided participants with the opportunity to make automation bias errors. The potential severity posed by these errors was such that they should be avoided in all circumstances. While there is little consensus regarding the assessment of severity,[2] many methods exist. Two methods with good validity and reliability are the NCC MERP,[3] originally developed to assess actual harm from errors, and a tool by Barber and Dean,[4] which requires at least four reviewers. The errors in the experiment were assessed using the method by Dornan and his colleagues.[5] It is a commonly used method, provides clear guidance on rating the potential severity of prescribing errors that occur independently of actual harm, and could be implemented by a single clinical pharmacist.

While automation bias has been extensively studied within multitask settings, clinical environments are also prone to interruptions, requiring clinicians to switch between tasks. Interruptions have the potential to increase cognitive workload and overall complexity.[6] However, the effect of interruption on automation bias has not yet been tested. Therefore, a secondary aim of the experiment was to test the impact of interruptions on automation bias.

3.2 Contribution of this article to thesis

The journal article (Article II) presented in this chapter seeks to address aims 2 to 4 of the thesis. It evaluates whether: (1) there is a risk of automation bias in electronic prescribing assisted by clinical decision support (aim 2); (2) high-complexity tasks are more susceptible to automation bias errors (aim 3); and (3) task interruptions increase the rate of automation bias errors (aim 4).

This article reports the results of the experiment on errors and finds evidence of automation bias omission and commission errors. However, task complexity and interruptions did not affect errors.

3.3 Article details

This article was published in *BMC Medical Informatics and Decision-making*.[7]

Citation

Lyell D, Magrabi F, Raban MZ, Pont LG, Baysari MT, Day RO, Coiera E: Automation bias in electronic prescribing. *BMC Medical Informatics and Decision-making* 2017, **17**(1):28.

The version of record is available from the publisher's website:

https://doi.org/10.1186/s12911-017-0425-5

3.4 Author contributions

David Lyell conceived this research and designed and conducted the study with guidance from, and under the supervision of, Enrico Coiera and Farah Magrabi. Magdalena Z. Raban designed the prescribing scenarios, with review and final approval by Richard O. Day. L.G. Pont designed the interruption tasks, scored the severity of prescribing errors contained in scenarios and the severity of errors made by participants. Melissa T. Baysari provided advice on the experiment design and the design of the simulated e-prescribing system. **David Lyell** drafted the manuscript with input from all authors. All authors provided revisions for intellectual content. All authors have read and approved the final manuscript.

Article II: Automation bias in electronic prescribing

Lyell et al. BMC Medical Informatics and Decision Making (2017) 17:28
DOI 10.1186/s12911-017-0425-5

BMC Medical Informatics and Decision Making

RESEARCH ARTICLE Open Access

Automation bias in electronic prescribing

David Lyell[1*], Farah Magrabi[1], Magdalena Z. Raban[2], L.G. Pont[2], Melissa T. Baysari[2,3], Richard O. Day[4] and Enrico Coiera[1]

Abstract

Background: Clinical decision support (CDS) in e-prescribing can improve safety by alerting potential errors, but introduces new sources of risk. Automation bias (AB) occurs when users over-rely on CDS, reducing vigilance in information seeking and processing. Evidence of AB has been found in other clinical tasks, but has not yet been tested with e-prescribing. This study tests for the presence of AB in e-prescribing and the impact of task complexity and interruptions on AB.

Methods: One hundred and twenty students in the final two years of a medical degree prescribed medicines for nine clinical scenarios using a simulated e-prescribing system. Quality of CDS (correct, incorrect and no CDS) and task complexity (low, low + interruption and high) were varied between conditions. Omission errors (failure to detect prescribing errors) and commission errors (acceptance of false positive alerts) were measured.

Results: Compared to scenarios with no CDS, correct CDS reduced omission errors by 38.3% ($p < .0001$, $n = 120$), 46.6% ($p < .0001$, $n = 70$), and 39.2% ($p < .0001$, $n = 120$) for low, low + interrupt and high complexity scenarios respectively. Incorrect CDS increased omission errors by 33.3% ($p < .0001$, $n = 120$), 24.5% ($p < .009$, $n = 82$), and 26.7% ($p < .0001$, $n = 120$). Participants made commission errors, 65.8% ($p < .0001$, $n = 120$), 53.5% ($p < .0001$, $n = 82$), and 51.7% ($p < .0001$, $n = 120$). Task complexity and interruptions had no impact on AB.

Conclusions: This study found evidence of AB omission and commission errors in e-prescribing. Verification of CDS alerts is key to avoiding AB errors. However, interventions focused on this have had limited success to date. Clinicians should remain vigilant to the risks of CDS failures and verify CDS.

Keywords: Decision support systems, Clinical, Cognitive biases, Complexity, Electronic prescribing, Medication errors, Automation bias, Human-computer interaction, Human-automation interaction

Background

The electronic prescription of medicines (e-prescribing) is now routine, [1] making the clinical decision support (CDS) systems they include [2] amongst the most common encountered by clinicians. CDS can help reduce adverse events by displaying alerts for potential errors such as drug-drug interactions [3–5].

However, CDS is not perfectly accurate and will at times provide inaccurate advice [6]. Over-reliance on alerts may cause clinicians to avoid prescribing particular medicines due to inappropriate alerts or clinicians may fail to detect prescribing errors with the potential for harm because they were not alerted to them.

This over-reliance on CDS is referred to as automation bias (AB), and is defined as "the tendency to use automated cues (such as CDS alerts) as a heuristic replacement for vigilant information seeking and processing [7]." With AB omission errors, users fail to notice problems because they were not alerted to the problem by CDS, and with commission errors, users comply with incorrect recommendations [7]. There are multiple possible causes of AB, [8, 9] and the literature is currently unclear regarding which, or all, of these are genuinely causal, and under which circumstances. For example, commission errors have been associated with reduced sampling of information which can verify decision support [10, 11]. However, human factors studies have found that some individuals make commission errors *despite* sampling all required information [12, 13]. This has been

* Correspondence: david.lyell@mq.edu.au
[1]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia
Full list of author information is available at the end of the article

described as a 'looking but not seeing' effect, suggesting that human information processing is also a factor affecting AB.

The majority of AB research comes from the human factors and ergonomics literature, mostly focused on aviation and process control [14]. There have been a small number of studies conducted in healthcare, finding evidence of AB omission errors in computer-aided detection of cancers in mammograms, [15, 16] and commission errors in the computerized interpretation of EKGs, [17] and answering questions about clinical scenarios [18]. Goddard, et al. [19] found evidence of commission errors, where general practitioners answered questions about which drugs they would prescribe for different clinical scenarios. They found a significant effect for participants changing from correct to incorrect responses after being provided with incorrect CDS advice.

For e-prescribing systems, decision support is commonly provided in the form of alerts that warn clinicians about potential prescribing errors [2]. Despite such alerts being one of the most common forms of decision support, the high volume of prescriptions ordered, and risk of harm to patients from prescribing errors, no studies have yet assessed the risk of AB in e-prescribing.

The prevailing view in the human factors literature is that AB only occurs in a multi-task environment [14, 20, 21]. However AB has been reported in some, but not all, tasks in a single task environment [14]. The discrepancy between single tasks which do and do not produce AB suggests that properties of the task itself may be risk factors for AB. The occurrence of AB may be related to how complex it is to verify that automation is working correctly, and that complexity across multiple simultaneous tasks appears to be cumulative [14]. In addition to multitasking, clinical settings are very prone to interruptions, requiring the clinician to switch between their primary task and the interruption, introducing increased cognitive workload and task complexity [22]. However, to date, no studies have tested the impact of interruptions on AB.

This study seeks to test for the presence of AB in e-prescribing assisted by CDS, which provides decision support in the form of alerts for prescribing errors. Additionally, it seeks to test the impact of interruptions and task complexity on AB. In doing so we seek to understand: (1) The baseline impact of correct CDS alerts on prescribing errors; (2) The impact of CDS false negatives on omission errors; (3) The impact of CDS false positive alerts on commission errors; (4) The impact of interruptions on AB; (5) The impact of task complexity on AB.

Methods

Participants

One hundred and twenty students enrolled in the final two years of a medical degree at Australian universities participated in the study. Australian medical education uses an integrative approach where students learn patient and clinical content throughout their degree. By the final two years of their education, participants would have typically received training in rational and safe prescribing. They also complete the National Prescribing Curriculum, a series of online modules based on the prescribing principles outlined in the World Health Organisation's Guide to Good Prescribing [23]. Upon completion of these final two years, graduates would begin practice as junior medical officers.

Participants responded to advertisements emailed by medical schools or posted on social media via medical students' societies. Ethical approval was granted by the ethics committees of Macquarie University and the University of New South Wales. Participants were offered two movie vouchers and a certificate for their participation.

Experiment design

The study had two within-subject factors: quality of CDS (correct, incorrect and no CDS) and task complexity (low, low with interruption and high) providing nine conditions (Fig. 1). Each participant received all nine conditions, completing one scenario in each condition. The experimental control were scenarios presented to participants with no CDS.

The allocation of the nine prescribing scenarios to the nine experimental conditions, the order of presentation, and whether participants received control scenarios first or last were randomized. The position of prescribing and false positive errors in the list of medicines to be prescribed was randomised, allocated at the time of scenario design. The position of alerts was varied depending on the CDS condition that was randomly allocated to the scenario for each participant at the time of enrolment.

Experimental task

Figure 2 provides an example of the participants' task in this experiment. Participants were presented with nine prescribing scenarios for which they were asked to prescribe medicines using an e-prescribing system. Each scenario presented a brief patient history together with a list of medications to prescribe.

The prescribing scenarios were developed with advice from an expert panel, including four hospital doctors, a medical pharmacology registrar and two pharmacists (including MZR). They were independently reviewed by a consultant physician specialising in pharmacology (RD), to ensure clinical relevance. The scenarios presented

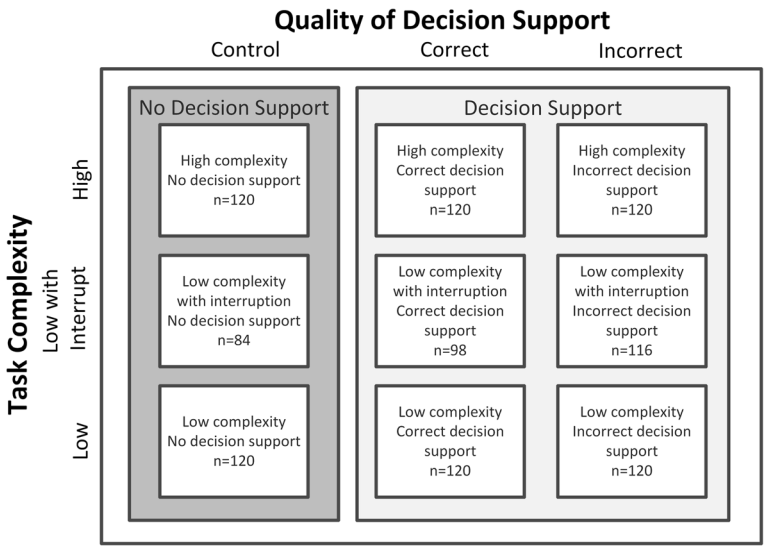

Fig. 1 Experimental design with the number of participants in each condition. All participants completed all conditions. However, some were excluded from the analysis of interruption conditions as they did not trigger the interruption task

hypothetical patient scenarios and involved prescribing tasks that were typical of those undertaken by junior medical officers, based on observations of e-prescribing in a medical ward of a major teaching hospital. A common task performed by junior medical officers is the prescribing of medications using an e-prescribing system upon admission of a patient to hospital, including medicines taken prior to, and those initiated on admission.

Each scenario included one genuine prescribing error, where one of the medicines was clinically contraindicated in that scenario (Additional file 1: Appendix A). These were designed to be unambiguously errors and of sufficient severity in the risk posed to the patient that the medicine should be avoided under all circumstances. To ensure this, the severity of the errors included in the scenarios were independently assessed by a clinical pharmacist (LGP). The error in one scenario was assessed as potentially lethal, five were serious, and three were significant [24]. All other medicines listed in scenarios where carefully chosen so as to be unambiguously free from error.

Scenario complexity was manipulated by varying the amount of information contained in the prescribing scenarios [25]. The nine scenarios were divided into six low-complexity scenarios (each containing six information elements) and three high complexity scenarios (each containing seventeen elements). An information element was classified as either a condition, symptom, test result, prior treatment, allergy, observation, or requested prescription. Each element could potentially interact with other elements in a way that could result in a prescribing error, for example, drug-drug interactions and conditions

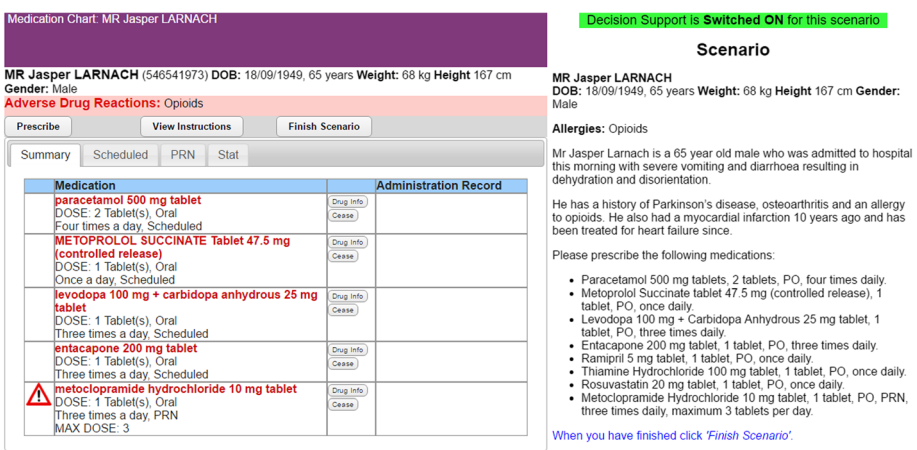

Fig. 2 The e-prescribing system interface and scenario

which may contraindicate the use of a particular medicine. The more elements, the more potential interactions the participant needs to assess. Low-complexity scenarios contained a list of three requested medicines, while high-complexity ones contained eight. The number of elements in each scenario was coded by DL and reviewed by MZR; disagreements were resolved by consensus.

In interruption conditions, participants were interrupted, once per scenario, whilst viewing drug information and presented with a task requiring a response before they could continue. The task required them to seek out and retain in memory three information elements to calculate a dose (Additional file 1: Appendix B).

E-prescribing system

A simulated e-prescribing system was developed which allowed for the manipulation of the triggering and content of CDS alerts. This web-based system was presented to participants as being in development. A medication administration record was not implemented, nor were participants required to specify times of administration.

CDS was provided in the form of alerts (Fig. 3) which were triggered once a prescription was entered. The alert provided a generic warning about the nature of the error, followed by specific details.

Participants could resolve the alert by choosing either to remove (i.e., not prescribe) the medicine or to override the alert with a reason and prescribe that medicine anyway. The alert also provided direct access to drug information for the relevant medicine from the Australian Medicines Handbook [26]. The Australian Medicines Handbook references the Australian formulary and is a gold standard medicines reference. It is evidence-based, reflects Australian best practice and is widely utilised in Australian clinical practice [27]. This reference was also readily accessible from the medication chart and in prescription order entry screens and could be used to identify prescribing errors and verify the information provided by CDS alerts.

The quality of CDS provided to participants was manipulated across conditions:

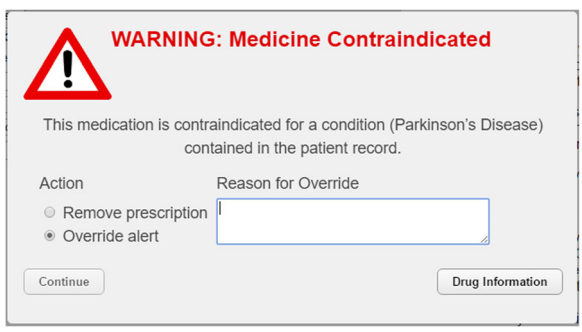

Fig. 3 CDS Alert

- *Correct CDS* alerts triggered only by genuine prescribing errors (true positives). Due to the severity of the prescribing errors, all correct alerts were highly relevant. The absence of alerts always indicated true negatives.
- *Incorrect CDS* failed to alert the genuine prescribing error (false negative) and provided one false positive alert, per scenario, for a medicine that was safe to prescribe.
- *No CDS* served as the control condition in which there was no CDS checking for errors. Participants were informed of this and advised to use the drug reference to identify errors.

Procedure

After having given informed consent, participants completed a pre-experiment questionnaire and watched a brief instructional video on how to use the e-prescribing system. The video included a demonstration of the correct functioning of CDS alerts and how to view drug information.

Participants were instructed as follows: (1) Approach tasks as if they were treating a real patient, exercising all due care; (2) Should they detect any prescribing errors, these should be addressed by not prescribing that medicine; (3) If the error involved an adverse drug interaction between two medicines, only one should be omitted; (4) If there was a discrepancy between CDS and the drug information they should rely on the drug information.

The task was presented as an evaluation of an e-prescribing system under development and participants were told that "Initial testing has shown that alerts are highly accurate, but occasionally have been incorrect. Therefore, you should always double check with the inbuilt drug information reference." No information was provided on what types of errors the system would check and alert. Once all scenarios were completed, participants completed a post-experiment questionnaire and were then debriefed.

Outcome measures

The present study was designed to test and analyze the following decision errors:

1. Omission errors: Where the participant failed to detect a genuine prescribing error. If the error was corrected by the participant, for example, by reducing a harmful dose to a safe level, it was not scored as an omission error.
2. Commission errors: Where the participant did not prescribe a safe medicine because of a false positive alert.

Prescribing errors were classified according to the definitions of prescribing error categories provided by Westbrook et al. [28]. The potential severity of prescribing errors was assessed by a clinical pharmacist (LGP) using the severity error classification scheme described in Dornan et al. [24].

Statistical analyses

The presence of AB was tested using McNemar's test [29] comparing errors between scenarios with incorrect CDS and scenarios with no CDS (control). It was estimated that 120 participants would be required to detect a 25% or greater difference (two-tailed) in errors between the control and incorrect CDS scenarios with 80% power and $p < 0.05$ [30]. With five hypotheses tested, a Bonferroni correction was applied to control for the increased risk of making a Type I error when testing multiple hypotheses [31]. With the desired alpha of 0.05, the corrected alpha against which all significance probabilities were evaluated became 0.01. Significance probabilities are only reported for comparisons between individual conditions, but not for aggregate figures by quality of CDS, which include multiple observations from each participant. Scenarios in which participants did not experience an interruption were excluded from the interruption analysis ($n = 36$ with no CDS, $n = 22$ with correct CDS, and $n = 4$ with incorrect CDS).

Results

The participants' average age was 24 years, and 46.7% were female. The majority rated their knowledge of medicines as fair (55.8%, $n = 67$) and only 5.8% ($n = 7$) reported previous training in e-prescribing systems. One participant completed the experiment twice (on two separate occasions), and the data from their second attempt was excluded.

In total, participants prescribed 4,065 medicines and made 1,049 prescribing errors (Table 1). This included 440 necessary medicines that were not prescribed. Of the total errors, 735 (70%) errors stemmed from opportunities the experiment provided for participants to make omission or commission errors. The remaining 314 (30%) where user-originated errors, independent of the experiment design and the majority of these were transcription errors. All participants made one or more prescribing errors. Compared to the control, correct CDS decreased prescribing errors by 58.8%, while incorrect CDS increased errors by 86.6%.

Although participants were instructed to omit medicines they believed to contain prescribing errors, there were 43 instances where it appeared participants had substituted medicines not included in the scenario in an attempt to correct errors. Of these, 36 substitutions were replacing medicines associated with genuine prescribing errors, six were in response to false positive alerts, and one substitution was for a medicine not associated with any experimental manipulation.

Correct CDS decreased prescribing errors

There were 40.8% fewer omission errors in scenarios with correct CDS compared to scenarios with no CDS (Table 2). This was significant across all levels of task complexity, with 38.3% fewer errors in low complexity ($p < .0001$, $n = 120$), 46.6% fewer errors in low + interrupt ($p < .0001$, $n = 70$), and 39.2% fewer errors in high complexity scenarios ($p < .0001$, $n = 120$).

However, correct CDS did not significantly alter commission errors (Table 2). This was the case for low complexity ($p = 1.0$, $n = 120$), low + interrupt ($p = .219$, $n = 65$) and high complexity scenarios ($p = .678$, $n = 120$). Participants also made omission errors by overriding correct CDS alerts in 8.3% of scenarios and commission errors by not prescribing the safe, comparator, medicines in 5.3% of scenarios.

Incorrect CDS increased prescribing errors

Participants missed 28.7% more genuine prescribing errors (omission errors) when assisted by incorrect CDS compared to no CDS (Table 2). These differences were statistically significant across all levels of complexity, with 33.3% more errors in low complexity ($p < .0001$, $n = 120$), 24.5% more errors in low + interrupt ($p = .009$, $n = 82$) and 26.7% more errors in high complexity scenarios ($p < .0001$, $n = 120$).

Overall participants made 56.9% more commission errors (did not prescribe safe medicines) when they received false positive alerts from incorrect CDS compared to when they received no CDS (Table 2). These differences were statistically significant across all levels of complexity, with participants in scenarios receiving false positive alerts making 65.8% more errors in low complexity ($p < .0001$, $n = 120$), 53.5% more errors in low + interrupt ($p < .0001$, $n = 82$) and 51.7% more errors in high complexity scenarios ($p < .0001$, $n = 120$).

Interruptions to prescribing and scenario complexity did not impact automation bias

Interruptions did not affect omission or commission errors, or errors in the control scenarios. In interrupted scenarios with incorrect CDS there were 0.1% more omission errors ($p = 1.0$ $n = 116$) and 9.7% fewer commission errors ($p = .08$, $n = 116$). In interrupted control scenarios there were 8.9% more omission errors ($p = .2$, $n = 84$) and 2.6% more commission errors ($p = .22$, $n = 84$). All of these were non-significant.

Scenario complexity did not affect omission or commission errors, or errors in the control scenarios. In high complexity scenarios with incorrect CDS there were 4.2% fewer omission errors ($p = .46$, $n = 120$) and 5.0%

Table 1 Prescribing errors

	Control		Quality of Decision Support					
	No CDS		Correct CDS		Incorrect CDS		Total	
	n	%	n	%	n	%	n	%
Omission errors								
Wrong drug	57	35.8	7	25.0	116	41.9	180	38.8
Wrong dose	55	34.6	9	32.1	72	26.0	136	29.3
Wrong frequency	0	0.0	0	0.0	0	0.0	0	0.0
Drug-drug interaction	28	17.6	7	25.0	60	21.7	95	20.5
Wrong route	0	0.0	0	0.0	0	0.0	0	0.0
Wrong formulation	0	0.0	0	0.0	0	0.0	0	0.0
Duplicated drug therapy	19	11.9	5	17.9	29	10.5	53	11.4
Not indicated	0	0.0	0	0.0	0	0.0	0	0.0
Not Prescribed	0	0.0	0	0.0	0	0.0	0	0.0
Total omission errors	159		28		277		464	
Commission errors								
Not prescribed	24	100.0	18	100.0	229	100.0	271	100.0
Total commission errors	24		18		229		271	
User originated errors								
Wrong drug	8	5.8	10	11.6	9	9.9	27	8.6
Wrong dose	43	31.4	29	33.7	22	24.2	94	29.9
Wrong frequency	10	7.3	6	7.0	4	4.4	20	6.4
Drug-drug interaction	0	0.0	0	0.0	0	0.0	0	0.0
Wrong route	0	0.0	0	0.0	1	1.1	1	0.3
Wrong formulation	1	0.7	0	0.0	0	0.0	1	0.3
Duplicated drug therapy	0	0.0	0	0.0	1	1.1	1	0.3
Not indicated	0	0.0	1	1.2	0	0.0	1	0.3
Not prescribed	75	54.7	40	46.5	54	59.3	169	53.8
Total user originated errors	137		86		91		314	
Total errors								
Wrong drug	65	20.3	17	12.9	125	20.9	207	19.7
Wrong dose	98	30.6	38	28.8	94	15.7	230	21.9
Wrong frequency	10	3.1	6	4.5	4	0.7	20	1.9
Drug-drug interaction	28	8.8	7	5.3	60	10.1	95	9.1
Wrong route	0	0.0	0	0.0	1	0.2	1	0.1
Wrong formulation	1	0.3	0	0.0	0	0.0	1	0.1
Duplicated drug therapy	19	5.9	5	3.8	30	5.0	54	5.1
Not indicated	0	0.0	1	0.8	0	0.0	1	0.1
Not Prescribed	99	30.9	58	43.9	283	47.4	440	41.9
Total errors	320		132		597		1049	

fewer commission errors ($p = .35$, $n = 120$). In high complexity control scenarios there were 2.5% more omission errors ($p = .75$, $n = 120$) and 9.2% more commission errors ($p = .007$, $n = 120$). The only significant difference was between low and high complexity control scenarios.

More omission than commission errors

Overall participants made 13.5% more omission than commission errors when provided with incorrect CDS, however, this was only significant in the low complexity + interrupt condition, all others were non-significant. There

Table 2 Number of participants making omission and commission errors

Scenario complexity	Quality of Decision Support											
	Control (No CDS)				Correct CDS				Incorrect CDS			
	Omission No alert ($n = 120$)		Commission No alert ($n = 120$)		Omission True positive alert ($n = 120$)		Commission True negative alert ($n = 120$)		Omission False negative alert ($n = 120$)		Commission False positive alert ($n = 120$)	
	n	%	n	%	n	%	n	%	n	%	n	%
Low	55	45.8	4	3.3	9	7.5	5	4.2	95	79.2	83	69.2
Low + Interrupt[a]	46	54.8	5	6	8	8.2	1	1	92	79.3	69	59.5
High	58	48.3	15	12.5	11	9.2	12	10	90	75	77	64.2
Total	159	49.1	24	7.4	28	8.3	18	5.3	277	77.8	229	64.3

[a]Number of participants in low + interrupt conditions: Control ($n = 84$), Correct DS ($n = 98$), and Incorrect DS ($n = 116$)

were 10% more omission errors in low complexity ($p = .065$, $n = 120$), 19.8% more in low + interrupt ($p = .001$, $n = 116$) and 10.8% more in high complexity scenarios ($p = .079$, $n = 120$).

Discussion

Main findings

This is the first study to find evidence of automation bias in the presence of e-prescribing CDS alerts. We found that when CDS was correct it reduced overall prescribing errors by 58.8%. This is consistent with prior literature showing that e-prescribing CDS can reduce prescribing errors [3–5]. However, when CDS was incorrect it increased errors by 86.6%. This increase was due to AB, that is, the ability of incorrect CDS to adversely influence participant prescribing decisions.

We found evidence of participants making omission errors, by failing to detect 28.7% more prescribing errors when CDS failed to provide alerts, compared to a control condition with no CDS. This finding was significant across all levels of task complexity and is potentially serious as the missed prescribing errors were classified as being of significant to potentially lethal severity, with most classified as serious severity.

Likewise, participants made commission errors, acting on clinically incorrect, false positive alerts, by not prescribing 56.9% more necessary medicines compared to the control condition. This was significant across all levels of task complexity.

These findings are consistent with and add to the research on automation bias in healthcare. Finding evidence of omission errors in the computer-aided detection of cancers in screening mammography [15, 16] and commission errors in the computerized interpretation of EKGs, [17] answering clinical questions assisted by CDS, [18] and deciding what to prescribe for clinical scenarios [19].

Interestingly, while participants were found to overrely on automation, there was evidence of disagreement with the CDS provided to them. Participants' overrode correct alerts and in doing so made prescribing errors which CDS was warning them to avoid. They also did not prescribe medicines which did not contain errors and for which there were no alerts. Reasons provided for overriding correct CDS alerts commonly referred to the condition for which the medicine was intended to treat (e.g. "VTE risk and pain management", "vomiting") or indicated that the medicine was regularly taken by the patient (e.g. "patient usual dose"). Participants commonly cited the lack of a true contraindication as the reason for overriding incorrect CDS alerts with many referring to the drug information. For example, "There is not any interaction listed on the drug information". However, regular patient medicines and the condition treated were also mentioned as reasons for overriding incorrect CDS alerts. This suggests that not only did participants have trouble determining when CDS was wrong, but some also had trouble recognizing when it was right and that the alerts, or lack thereof, were beneficial and should be heeded.

Interruptions and task complexity did not impact automation bias

Interruptions did not affect the rate of AB errors nor did it affect errors rates in the control condition. However, interruptions are a complex phenomenon where multiple variables, including the characteristics of primary tasks, an individual's cognitive state, the interruptions themselves, and the environment, may influence impact on clinical tasks and errors [22]. Despite clear evidence that interruptions can disrupt clinical tasks, their effects are complex, and may not always be detected [32].

Any impact of interruptions on prescribing errors was not detected in our experiment, replicating earlier results [33]. In our experiment, upon task resumption participants had ample time to recall their next action and the task environment provided cues to aid task resumption, for example, partly completed orders were visible on screen. One possible reason for not detecting an effect of interruptions was thus that disruption were minimized by these cues within the user interface [34]. This is consistent

with observations from other studies of interruptions to computer-based tasks where participants were aided by the screen environment and were able to resume an interrupted task [35, 36]. Performance under cognitive load from more demanding competing tasks in a clinical environment may have resulted in a different outcome.

Contrary to expectations, the task complexity manipulation also had no effect on AB errors. This is in stark contrast to the findings of Bailey and Scerbo [25] who found performance on a system monitoring task deteriorated with increased task complexity, which they defined in terms of the cognitive demands placed on the participant. Monitoring tasks required the identification of critical deviations outside the normal operating range. Less complex tasks had participants monitor analogue gauges with marked critical regions. More complex tasks involved monitoring a display showing raw numbers where the subject had to remember the critical values for four different types of parameters.

Had the complexity manipulation altered the difficulty of prescribing task we would have expected to see a higher error rate in the high complexity control conditions. However, the observed difference was small and non-significant. This is in contrast to findings of Goddard et al. [19]. who found a significant effect for task difficulty, as classified by a panel of practitioners, on decision accuracy without CDS between medium and difficult scenarios. However, they found that task difficulty had no effect on commission errors.

The high error rate for both high and low levels of complexity in control conditions, with participants missing nearly half of all prescribing errors, seems to indicate that the difference in complexity between the two conditions may not have been large enough for differences in error rates to emerge.

Implications

When clinical decision support is right, it can reduce prescribing errors by providing an important opportunity to detect and recover from prescribing errors. However, the finding of automation bias suggests that this additional layer of defence weakens or, at worst, becomes a replacement for the clinician's own efforts in error detection with error detection delegated to CDS, without adequate oversight.

An intuitive solution to the problem of AB is to produce CDS systems that are less prone to error. While this may reduce the overall error rate, highly accurate automation is known to increase the rate of AB [25]. In other words, when automation does fail, the clinician will be even less able to detect it.

A key problem is that users seem to have difficulty in determining when CDS should and should not be relied on. Indeed, human factors research reports an inverse relationship between measures of verification, such as viewing drug references, and AB commission errors [10, 11]. So far, interventions to counter AB have had little success [37–39]. These include a number specifically targeted at verification, such as exposure to automation failures; [10] training about AB; and providing prompts to verify [40]. Compounding this problem further are findings of a *looking-but-not-seeing* effect or *inattentional blindness* where participants have made AB errors despite accessing sufficient information to assess that automation was incorrect [12, 13].

Verification, the means by which a user can determine whether the CDS they receive is correct, is key to the mitigation of AB. However, the lack of successful interventions indicates that more research is needed on how to best assist users with this crucial task.

This study has established that there is a risk of automation bias in electronic prescribing with senior medical students, who will soon be entering clinical practice as junior doctors. In doing this, we have also demonstrated a methodology for detecting AB in e-prescribing. The true rates and effects of AB in working clinical settings will require further studies and indeed is likely to vary by clinicians' experience and familiarity with medications, clinical setting, patient complexity, and the particular decision support system used. All this is future work. Likewise, the lack of an effect of task complexity, even in control conditions, was surprising and something future studies will need to address. This might be achieved by varying clinician experience with prescribing and e-prescribing systems. Complexity could also incorporate familiarity with medicines, varying between simple, commonly-used to complex, rarely-prescribed regimes.

Clinicians need to be mindful that CDS can and does fail [6]. Ideally, clinicians should make every effort to detect prescribing errors, allowing CDS to function as an independent check for errors rather than relying on it as a replacement of their own error detection efforts.

Limitations

Several limitations arise from the design of this study. While participants were instructed to approach the task as if they were treating a real patient, exercising all due care, the prescribing task was simulated, and prescribing errors were without consequence.

Also as an experiment, we cannot make any inferences about the true effect size or rate of AB in clinical settings as this will vary with, the user, the tasks being performed and the accuracy of the decision support provided. Likewise, the nature and incidence of the provided opportunities for prescribing errors may not be representative of those encountered in clinical practice.

The lack of a difference in prescribing errors between the low and high complexity control scenarios limited our ability to assess the impact of task complexity on AB.

Finally, the use of medical students with little experience in both prescribing medicines and using e-prescribing systems provides an indication of how CDS will impact new clinicians entering practice but limits generalisability for experienced prescribers or clinicians with e-prescribing experience.

Conclusion

This study set out to test for the presence of automation bias in e-prescribing, a clinical decision support system commonly encountered by clinicians. We found evidence of omission errors, where participants failed to detect prescribing errors that were not alerted by CDS and commission errors, where participants acted on clinically incorrect alerts. Contrary to expectations, task complexity and interruptions had no impact on AB errors. However, when prescribing errors were correctly alerted, there was a dramatic reduction in the number of prescribing errors, demonstrating the benefits of CDS.

The challenge is to maximize the benefits of CDS while minimizing the risk of over-reliance. The key to this is enabling clinicians to determine when the CDS provided to them is correct, which is achieved through verification. Unfortunately, interventions tested to date, including those which focus on verification have produced little success. More research is needed on how to best assist clinicians with the task of verifying automation.

Additional file

Additional file 1: Appendix A Overview of prescribing scenarios and Appendix B Example of an interruption task. (PDF 113 kb)

Abbreviations

AB: Automation Bias; CDS: Clinical Decision Support; EKG: Electrocardiogram; e-prescribing: Electronic Prescribing

Acknowledgments

Vitaliy Kim developed the e-prescribing system.
Jingbo Liu provided additional programming support.
Robin Butterfield reviewed and provided feedback on the prescribing scenarios.
Monish Maharaj pilot tested the experiment providing critical feedback on instruction to participants and advice on participant recruitment.

Funding

This research was supported by a Doctoral Scholarship for David Lyell provided by the HCF Research Foundation.

Availability of data and materials

Not applicable.

Authors' contributions

DL conceived this research, and designed and conducted the study with guidance and under the supervision of EC and FM. MZR designed the prescribing scenarios, with review and final approval by RD. LGP designed the interruption tasks, scored the severity of prescribing errors contained in scenarios and the severity of errors made by participants. MB provided advice on the experiment design and the design of the simulated e-prescribing system. DL drafted the manuscript with input from all authors. All authors provided revisions for intellectual content. All authors have read and approved the final manuscript.

Competing interests

The authors have no competing interests to declare.

Consent for publication

No personally identifying information is reported in this article.
Patients presented in the prescribing scenarios are fictional. All biographical information (including that depicted in Fig. 2) was made up for the purpose of this experiment in order to present participants with the information they would expect in such patient cases.

Ethics approval and consent to participate

This study received ethical approval from Macquarie University Human Research Ethics Committee (5201401029) and University of New South Wales Human Research Ethics Advisory Panel (2014-7-32).
Participant consent was obtained by an online participant information statement and consent form in accordance with approved protocols. At the conclusion of the experiment, participants were fully debriefed.

Publisher's Note

Author details

[1]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia. [2]Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia. [3]St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Sydney, Australia. [4]St Vincent's Hospital Clinical School and Pharmacology, School of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney, Australia.

Received: 25 November 2016 Accepted: 9 March 2017
Published online: 16 March 2017

References

1. Britt H, Miller G, Henderson J, Bayram C, Harrison C, Valenti L, Wong C, Gordon J, Pollack A, Pan Y, et al. General practice activity in Australia 2014–15. General practice series no. 38. Sydney: Sydney University Press; 2015.
2. Sweidan M, Williamson M, Reeve JF, Harvey K, O'Neill JA, Schattner P, Snowdon T. Evaluation of features to support safety and quality in general practice clinical software. BMC Med Inform Decis Mak. 2011;11(1):1–8.
3. Wolfstadt JI, Gurwitz JH, Field TS, Lee M, Kalkar S, Wu W, Rochon PA. The effect of computerized physician order entry with clinical decision support on the rates of adverse drug events: a systematic review. J Gen Intern Med. 2008;23(4):451–8.
4. Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The effect of electronic prescribing on medication errors and adverse drug events: a systematic review. J Am Med Inform Assoc. 2008;15(5):585–600.
5. van Rosse F, Maat B, Rademaker CMA, van Vught AJ, Egberts ACG, Bollen CW. The effect of computerized physician order entry on medication prescription errors and clinical outcome in pediatric and intensive care: a systematic review. Pediatrics. 2009;123(4):1184–90.
6. Wright A, Hickman T-TT, McEvoy D, Aaron S, Ai A, Andersen JM, Hussain S, Ramoni R, Fiskio J, Sittig DF, et al. Analysis of clinical decision support system malfunctions: a case series and survey. J Am Med Inform Assoc. 2016;23(6):1068–76.
7. Mosier KL, Skitka LJ. Human decision makers and automated decision aids: made for each other. In: Parasuraman R, Mouloua M, Hillsdale NJ, editors. Automation and human performance: theory and applications. England: Lawrence Erlbaum Associates; 1996. p. 201–20.
8. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012; 19(1):121–7.
9. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. Hum Factors. 2010;52(3):381–410.

10. Bahner J, Huper A-D, Manzey D. Misuse of automated decision aids: complacency, automation bias and the impact of training experience. Int J Hum Comput Stud. 2008;66(9):688–99.
11. Bahner JE, Elepfandt MF, Manzey D. Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias. In: Proceedings of the human factors and ergonomics society annual meeting: Sep 22–26 2008. New York: Human Factors And Ergonomics Society; 2008. p. 1330–4.
12. Reichenbach J, Onnasch L, Manzey D. Misuse of automation: the impact of system experience on complacency and automation bias in interaction with automated aids. In: Proceedings of the human factors and ergonomics society annual meeting: Sep 27 - Oct 1 2010. San Francisco: Human Factors And Ergonomics Society; 2010. p. 374–8.
13. Manzey D, Reichenbach J, Onnasch L. Human performance consequences of automated decision aids: the impact of degree of automation and system experience. J Cogn Eng Decis Mak. 2012;6(1):57–87.
14. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. J Am Med Inform Assoc. 2017;24(2):423–31.
15. Alberdi E, Povykalo A, Strigini L, Ayton P. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. Acad Radiol. 2004;11(8):909–18.
16. Povyakalo AA, Alberdi E, Strigini L, Ayton P. How to discriminate between computer-aided and computer-hindered decisions: a case study in mammography. Med Decis Making. 2013;33(1):98–107.
17. Bogun F, Anh D, Kalahasty G, Wissner E, Bou Serhal C, Bazzi R, Weaver WD, Schuger C. Misdiagnosis of atrial fibrillation and its clinical consequences. Am J Med. 2004;117(9):636–42.
18. Golchin K, Roudsari A. Study of the effects of clinical decision support system's incorrect advice and clinical case difficulty on users' decision making accuracy. Stud Health Tech Informat. 2011;164:13–6.
19. Goddard K, Roudsari A, Wyatt JC. Automation bias: empirical results assessing influencing factors. Int J Med Informat. 2014;83(5):368–75.
20. Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced "complacency.". Int J Aviat Psychol. 1993;3(1):1–23.
21. Molloy R, Parasuraman R. Monitoring automation failures: effects of automation reliability and task complexity. In: Proceedings of the human factors and ergonomics society annual meeting: Oct 12–16 1992. Atlanta, GA, USA: Human Factors And Ergonomics Society; 1992. p. 1518–21.
22. Li SYW, Magrabi F, Coiera E. A systematic review of the psychological literature on interruption and its patient safety implications. J Am Med Inform Assoc. 2011;19(1):6–12.
23. De Vries TPGM, Henning RH, Hogerzeil HV, Fresle DA. Guide to good prescribing. Geneva: World Health Organization; 1994.
24. Dornan T, Ashcroft D, Heathfield H, Lewis P, Miles J, Taylor D, Tully M, Wass V. An in-depth investigation into causes of prescribing errors by foundation trainees in relation to their medical education: EQUIP study. London: General Medical Council; 2009. p. 1–215.
25. Bailey NR, Scerbo MW. Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. Theor Issues Ergon Sci. 2007;8(4):321–48.
26. Australian Medicines Handbook 2015 (online). Adelaide: Australian Medicines Handbook Pty Ltd; 2015 January. Available from: http://amhonline.amh.net.au/.
27. Day RO, Snowden L. Where to find information about drugs. Aust Prescr. 2016;39(3):88–95.
28. Westbrook JI, Reckmann M, Li L, Runciman WB, Burke R, Lo C, Baysari MT, Braithwaite J, Day RO. Effects of two commercial electronic prescribing systems on prescribing error rates in hospital in-patients: a before and after study. PLoS Med. 2012;9(1):e1001164.
29. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika. 1947;12(2):153–7.
30. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39(2):175–91.
31. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ. 1995;310(6973):170.
32. Magrabi F, Li S, Dunn A, Coiera E. Why is it so difficult to measure the effects of interruptions in healthcare? Stud Health Tech Informat. 2010;160:784–8.
33. Magrabi F, Li SYW, Day RO, Coiera E. Errors and electronic prescribing: a controlled laboratory study to examine task complexity and interruption effects. J Am Med Inform Assoc. 2010;17(5):575–83.
34. Coiera E. Technology, cognition and error. BMJ Qual Saf. 2015;24(7):417–22.
35. Dodhia RM, Dismukes RK. Interruptions create prospective memory tasks. Appl Cogn Psychol. 2009;23(1):73–89.
36. Speier C, Valacich JS, Vessey I. The influence of task interruption on individual decision making: an information overload perspective. Decis Sci J. 1999;30(2):337–60.
37. Skitka LJ, Mosier K, Burdick MD. Accountability and automation bias. Int J Hum Comput Stud. 2000;52(4):701–17.
38. Bagheri N, Jamieson GA: The impact of context-related reliability on automation failure detection and scanning behaviour. In: IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat No04CH37583): Oct 10–13 2004; 2004: 212–217 Vol. 1.
39. Singh AL, Tiwari T, Singh IL. Performance feedback, mental workload and monitoring efficiency. J Indian Acad Appl Psychol. 2010;36(1):151–8.
40. Mosier KL, Skitka LJ, Dunbar M, McDonnell L. Aircrews and automation bias: the advantages of teamwork? Int J Aviat Psychol. 2001;11(1):1–14.

Article II: Appendices

Appendix A: Overview of prescribing scenarios

Scenario	Error Type	Medication	Alert displayed to participants	Comment	Severity
A **Low complexity**	Prescribing Error	Digoxin 250 microgram tablet, 2 tablets, PO, three times a day.	WARNING: High Dose The entered dose is higher than the recommended maintenance dose range.	The elderly patient has atrial fibrillation which was controlled with Digoxin prior to admission. The dose requested by the scenario is a loading dose. The maintenance dose for an elderly patient is 62.5 to 125 micrograms once daily.	Serious
	False Positive	Lisinopril 5mg tablet, 1 tablet, PO, once daily.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Pulmonary Oedema) contained in the patient record.	Lisinopril is not contraindicated in patients with pulmonary oedema.	
B **Low complexity**	Prescribing Error	Spironolactone 25mg tablet, 1 tablet, PO, once daily.	WARNING: Medicine Contraindicated This medicine is contraindicated for a condition (Hyperkalaemia) contained in the patient record.	Patient has hyperkalaemia for which Spironolactone is contraindicated.	Serious
	False Positive	Augmentin Duo Forte (amoxycillin 875 mg + clavulanic acid 125 mg) tablet, 1 tablet, PO, twice daily.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Systolic Heart Failure) contained in the patient record.	Augmentin Duo Forte is not contraindicated in patients with heart failure.	
C **Low complexity**	Prescribing Error	Warfarin Sodium 2 mg tablet, 1 tablet, PO, once daily. and Ibuprofen 400 mg tablet, 1 tablet, PO, three times daily.	WARNING: Adverse Drug Interaction This medication has a listed adverse interaction with another already prescribed medication. There is an adverse drug interaction for Warfarin and Ibuprofen. This combination should be avoided.	Non-steroidal anti-inflammatory drugs (such as Ibuprofen) increase the risk of gastrointestinal bleeding in patients taking Warfarin. The combination should be avoided, especially as better analgesic options are available.	Significant
	False Positive	Atorvastatin 10 mg tablet, 1 tablet, PO, once daily.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Venous Thromboembolism) contained in the patient record.	Atorvastatin is not contraindicated in patients with venous thromboembolism.	
D **Low complexity**	Prescribing Error	Aspirin 300 mg tablet: effervescent, 3 tablets, PO, every 6 hours.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Peptic Ulcer Disease) contained in the patient record.	Patient has peptic ulcer disease with a history of bleeds for which aspirin increases the risk of gastrointestinal ulceration. There are better analgesic options.	Significant
	False Positive	Pantoprazole 40 mg tablet: enteric, 1 tablet, PO, once daily.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Severe Vomiting) contained in the patient record.	Pantoprazole is not contraindicated in patients with severe vomiting.	
E **Low complexity**	Prescribing Error	Loperamide Hydrochloride 2 mg capsule, 1 capsule, PO, PRN, every four hours, maximum 8 capsules per day.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Ulcerative Colitis) contained in the patient record.	Loperamide is contraindicated in patients with ulcerative colitis which poses a risk of toxic megacolon.	Serious
	False Positive	Mesalazine 500 mg tablet: enteric, 1 tablet, PO, three times daily. and Prednisolone 25 mg tablet, 1 tablet, PO, once daily.	WARNING: Adverse Drug Interaction This medication has a listed adverse interaction with another already prescribed medication. There is an adverse drug interaction for Mesalazine and Prednisolone. This combination should be avoided.	There is no documented adverse drug interaction for Mesalazine and Prednisolone.	

Scenario	Error Type	Medication	Alert displayed to participants	Comment	Severity
F **Low complexity**	Prescribing Error	Phenelzine 15 mg tablet, 1 tablet, PO, three times daily. and Tramadol Hydrochloride 50mg capsules, 2 capsules, PO, PRN, every six hours, maximum 8 capsules per day.	WARNING: Adverse Drug Interaction This medication has a listed adverse interaction with another already prescribed medication. There is an adverse drug interaction for Phenelzine and Tramadol hydrochloride. This combination should be avoided.	The combination of phenelzine and tramadol are contraindicated due to the possibility of causing serotonin toxicity.	Serious
	False Positive	Ramipril 5 mg tablet, 1 tablet, PO, once daily.	WARNING: Adverse Drug Reaction This patient has an Allergy or Adverse Drug Reaction recorded for this medication.	The patient is allergic to Sulfonamide. However Ramipril is not contraindicated for this allergy.	
G **High complexity**	Prescribing Error	Paracetamol 500 mg tablet, 2 tablets, PO, four times a day. and Panadeine Forte (Codeine Phosphate with Paracetamol Tablet 30 mg-500 mg) tablet, 2 tablets, PO, PRN, every four hours, maximum 8 tablets per day.	WARNING: High Dose / Duplicate Substance Both Paracetamol and Panadeine Forte (Codeine Phosphate 30mg with Paracetamol 500mg) contain the ingredient Paracetamol. The total Paracetamol entered is higher than the recommended dose range.	Prescribed together these two prescriptions provide for a combined maximum possible dose of 8 grams of paracetamol per day, double the maximum daily dose of 4 grams.	Significant
	False Positive	Ciprofloxacin 250 mg tablet, 1 tablet, PO, twice daily.	WARNING: Adverse Drug Reaction This patient has an Allergy or Adverse Drug Reaction recorded for this medication.	The patient is allergic to penicillin. Ciprofloxacin is an antibiotic, however, it is not contraindicated for allergy to penicillin.	
H **High complexity**	Prescribing Error	Methotrexate 2.5 mg tablets, 3 tablets, PO, once daily.	WARNING: High Dose The entered dose is higher than the recommended maintenance dose range	Patient has net onset rheumatoid arthritis. For treatment of rheumatoid arthritis, the loading dose of methotrexate is 7.5mg once weekly.	Potentially lethal
	False Positive	Paracetamol 500 mg tablet, 2 tablets, PO, PRN, every four hours, maximum 8 tablets per day.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Peptic Ulcer Disease) contained in the patient record.	Patient has newly diagnosed peptic ulcer disease, however, it is not a contraindication for paracetamol.	
I **High complexity**	Prescribing Error	Metoclopramide Hydrochloride 10 mg tablet, 1 tablet, PO, PRN, three times daily, maximum 3 tablets per day.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Parkinson's Disease) contained in the patient record.	Patient has a history of Parkinson's disease for which Metoclopramide is contraindicated as symptoms may worsen. The drug reference provides an alternative medicine as being preferred.	Serious
	False Positive	Entacapone 200 mg tablet, 1 tablet, PO, three times daily. and Rosuvastatin 20 mg tablet, 1 tablet, PO, once daily.	WARNING: Adverse Drug Interaction This medication has a listed adverse interaction with another already prescribed medication. There is an adverse drug interaction for Entacapone and Rosuvastatin. This combination should be avoided.	There is no documented adverse drug interaction for Entacapone and Rosuvastatin.	

Appendix B: Example of an interruption task

Harold O'Brien is about to commence Vancomycin for meningitis. Please refer to the test Creatinine Clearance in the test results provided and select the appropriate dose and frequency of administration using the dosing tables below.

Clinical Chemistry		
Sodium (mmol/L)	**140**	135-145
Potassium (mmol/L)	**4.5**	3.5-5.0
Chloride (mmol/L)	**106**	95-107
Bicarbonate (mmol/L)	**28**	24-32
Urea (mmol/L)	**9.2**	3.0-8.0
Creatinine (mmol/L)	**141**	60-110
Creatinine Clearance (mL/min)	**64**	97-137

Creatinine clearance (mL/min)	**Starting maintenance dosage**
more than 90	1.5 g
90 or less	1 g

Creatinine clearance (mL/min)	**Frequency**
more than 60	12-hourly
20 to less than 60	24-hourly
less than 20	48-hourly

Please review the information above. What *dose* of Vancomycin should be given and how *frequently* should it be administered?

(a) 1.5g, 12-hourly

(b) 1.5g, 24-hourly

(c) 1.5g, 48-hourly

(d) 1g, 12-hourly

(e) 1g, 24-hourly

(f) 1g, 48-hourly

Chapter 3 summary

3.5 Effect of task complexity and clinical decision support on errors

The effects models presented in the chapter summaries provide a visual summary of the experimental findings reported in the chapter. The models are presented as a causal network which illustrates the monotonic relationships observed between experimental variables. The links between variables show the evaluated relationships, while the link polarity indicates the monotonic relationship between variables; positively linked variables vary in the same direction (increase in one is linked to an increase in the other and vice-versa), while negatively linked variables vary in opposite directions (an increase in one is linked to a decrease in the other and vice-versa).

Causality can be inferred from this controlled experiment where an independent variable impacts a dependent variable. Conversely, causality cannot be inferred for associations between two dependent variables.

3.5.1 Omission errors

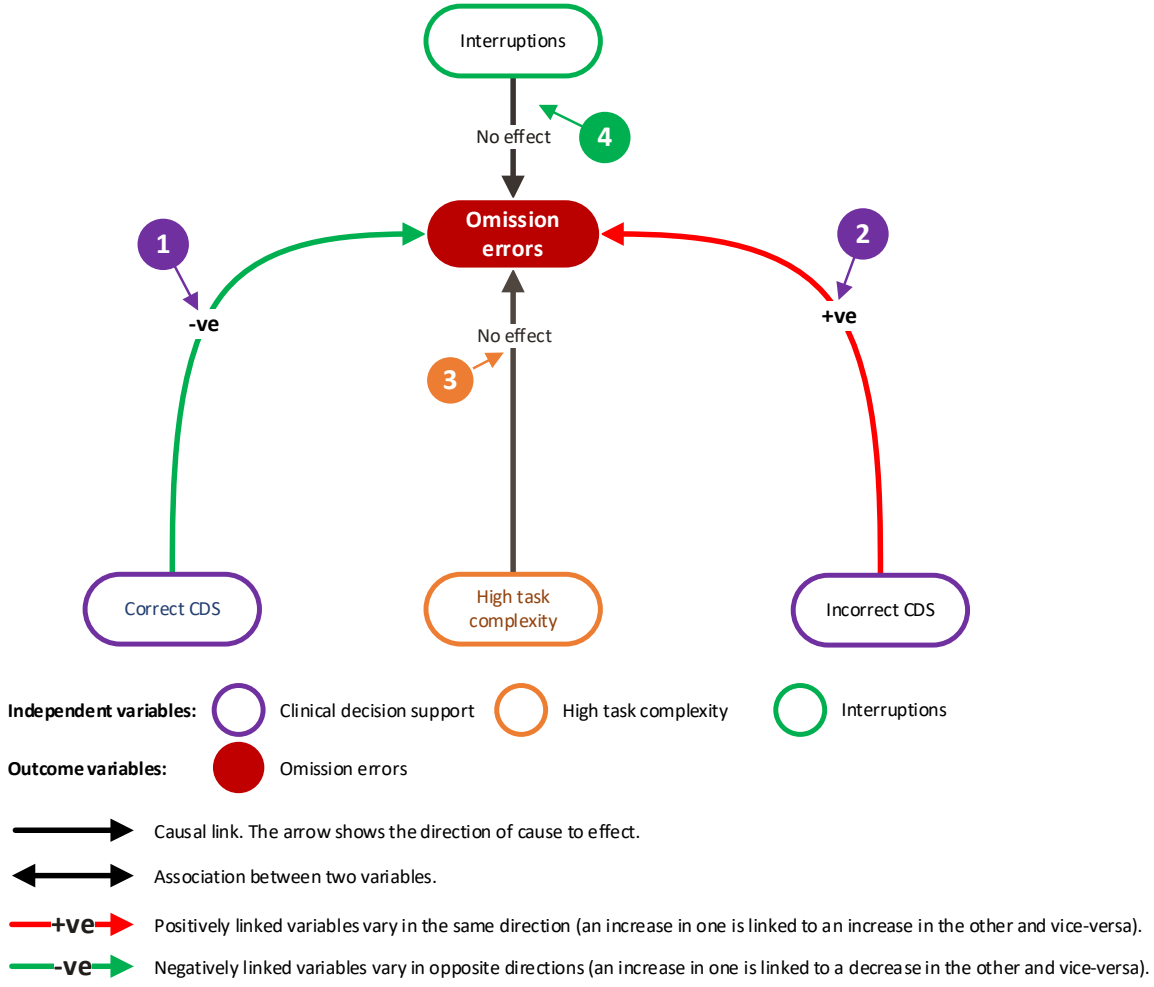

Figure 3-1 Effect of task complexity and clinical decision support on omission errors

1 Correct CDS decreases omission errors.

Correct CDS significantly reduced omission errors by 40.8% compared to when there was no CDS. See Table 2 in Article II.

2 Incorrect CDS increases omission errors.

Incorrect CDS significantly increased omission errors by 28.7% compared to when there was no CDS. See Table 2 in Article II.

3 High task complexity does not affect omission errors.

There were no significant differences in omission errors between low- and high-complexity with incorrect CDS or no CDS. See Table 2 in Article II.

4 Interruptions do not affect omission errors.

There were no significant differences in omission errors between low- and low-complexity with interruption scenarios with incorrect CDS or no CDS. See Table 2 in Article II.

3.5.2 Commission errors

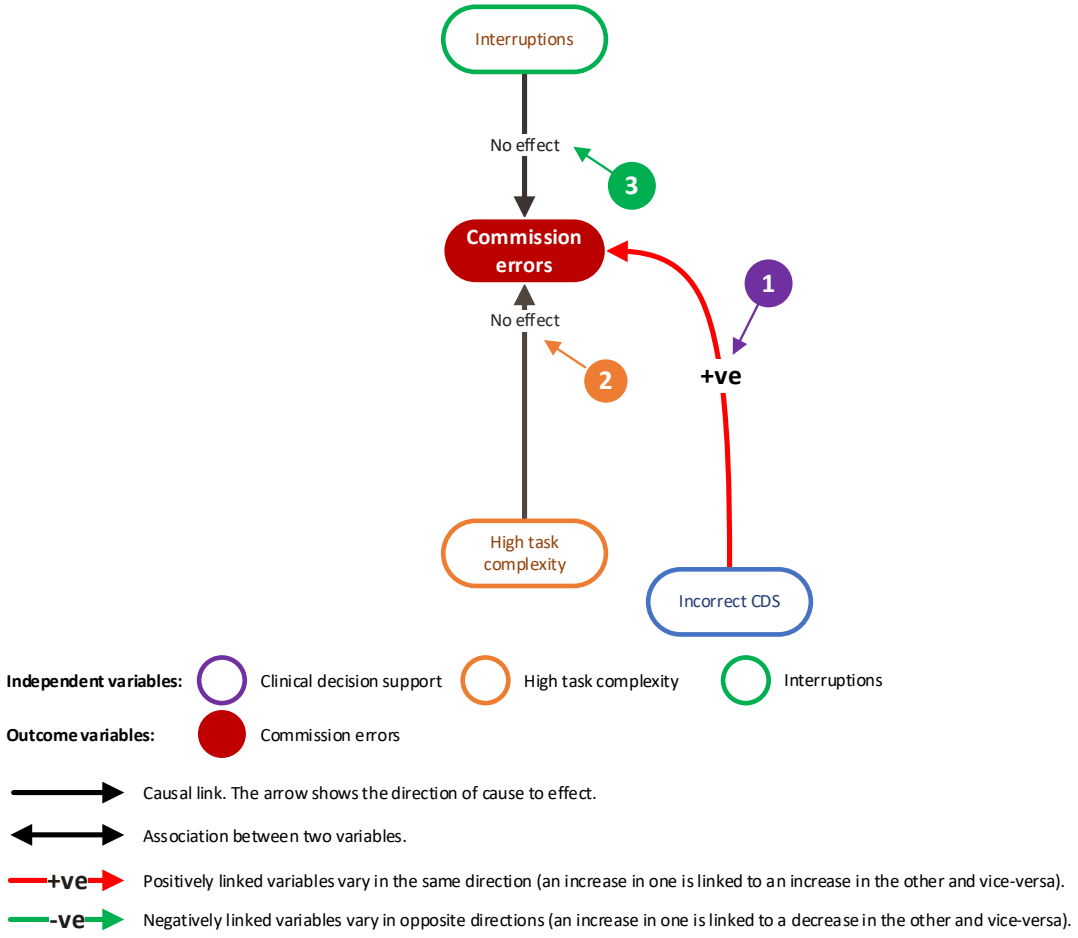

Figure 3-2 The effect of task complexity and clinical decision support on commission errors

1 Incorrect CDS increases commission errors.

Incorrect CDS significantly increased omission errors by 56.9% compared to when there was no CDS. See Table 2 in Article II.

2 High task complexity does not affect commission errors.

There was no significant difference between low- and high-complexity scenarios with incorrect CDS. See Table 2 in Article II.

3 Interruptions do not affect commission errors.

There was no significant difference between low- and low-complexity with interruption scenarios with incorrect CDS. See Table 2 in Article II.

3.6 Chapter 3 References

1. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association* 2017;**24**(2):423-31 doi: 10.1093/jamia/ocw105

2. Lewis PJ, Dornan T, Taylor D, Tully MP, Wass V, Ashcroft DM. Prevalence, Incidence and Nature of Prescribing Errors in Hospital Inpatients. *Drug Safety* 2009;**32**(5):379-89 doi: 10.2165/00002018-200932050-00002

3. Forrey RA, Pedersen CA, Schneider PJ. Interrater agreement with a standard scheme for classifying medication errors. *American Journal of Health-System Pharmacy* 2007;**64**(2):175-81 doi: 10.2146/ajhp060109

4. Barber ND, Dean BS. A validated, reliable method of scoring the severity of medication errors. *American Journal of Health-System Pharmacy* 1999;**56**(1):57-62 doi: 10.1093/ajhp/56.1.57

5. Dornan T, Ashcroft D, Heathfield H, Lewis P, Miles J, Taylor D, Tully M, Wass V. An in-depth investigation into causes of prescribing errors by foundation trainees in relation to their medical education: EQUIP study. *London: General Medical Council* 2009:1-215

6. Li SYW, Magrabi F, Coiera E. A systematic review of the psychological literature on interruption and its patient safety implications. *Journal of the American Medical Informatics Association* 2011;**19**(1):6-12 doi: 10.1136/amiajnl-2010-000024

7. Lyell D, Magrabi F, Raban MZ, Pont LG, Baysari MT, Day RO, Coiera E. Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making* 2017;**17**(1):28 doi: 10.1186/s12911-017-0425-5

4 Automation bias and cognitive load

This chapter examines the relationship between automation bias and information processing. Information processing is assessed in terms of the cognitive load experienced by participants as they performed experimental tasks.

4.1 Background

The systematic review [Chapter 2; 1] identified task complexity as a potential risk factor for automation bias in healthcare tasks and proposed Cognitive Load Theory as a framework for assessing its effects.

It was originally hypothesised that high task complexity would increase participants' cognitive load, which would lead to a greater reliance on decision support to manage the cognitive demands of the task. This, in turn, was hypothesised to lead to increased automation bias errors when decision support is incorrect. While Article II [Chapter 3; 2] reported that task complexity did not affect automation bias errors, an analysis of the cognitive load effects was undertaken to establish how it was affected by automation bias errors, task complexity, and the quality of clinical decision support. This analysis is reported in Article III [3] and is the focus of this chapter.

4.1.1 Exclusion of the interruption condition from further analyses

Article II [Chapter 3; 2] reported no effect of interruptions on errors; therefore, the interruption condition was excluded from further analyses. This was to firstly facilitate the concise presentation of the cognitive load and verification analyses within self-contained journal articles. Secondly, while interruptions may contribute to cognitive load,[4] they are nevertheless a complex phenomenon [4] which exceed the scope and aims of this thesis. Instead, further analyses will focus on the task complexity manipulation which was designed to impact participants' cognitive load.

4.2 Contribution of article to thesis

Article III [3] tests the hypothesis set out in aim 5: as task complexity increases, so too will participants' cognitive load, which, in turn, would lead to greater reliance on clinical decision support to manage increasing cognitive demands and prevent overload.

This article reports the results of the experiment on participants' self-reported cognitive load. It confirms that increasing task complexity from low to high did significantly increase participants' cognitive load. The major findings are that the presence of decision support lowered cognitive load in high complexity tasks. However, omission errors were significantly associated with lower cognitive load in both incorrect and no CDS conditions. This refutes the hypothesis set out in aim 5. Errors in the

control condition without decision support were associated with lower cognitive load, suggesting that errors did not result from cognitive overload, but may be due to insufficient cognitive resources being allocated to the task.

The findings reported in this article further contribute to an understanding of the relationship between automation bias errors, task complexity and cognitive load. This contributes to aim 6, which will be addressed in the discussion (chapter 6).

4.3 Article details

This article was published in *Human Factors: The Journal of the Human Factors and Ergonomics Society*.

Citation

Lyell D, Magrabi F, Coiera E. The Effect of Cognitive Load and Task Complexity on Automation Bias in Electronic Prescribing. *Human Factors* 2018;**60**(7):1008 - 21 doi: 10.1177/0018720818781224

The version of record is available from the publisher's website:

https://doi.org/10.1177/0018720818781224

4.4 Author contributions

David Lyell conceived this research and designed and conducted the study with guidance from, and under the supervision of, Enrico Coiera and Farah Magrabi.

David Lyell drafted the manuscript with input from all authors. All authors provided revisions for intellectual content. All authors have approved the final manuscript.

Article III: The effect of cognitive load and task complexity on automation bias in electronic prescribing

The Effect of Cognitive Load and Task Complexity on Automation Bias in Electronic Prescribing

David Lyell, **Farah Magrabi**, and **Enrico Coiera**, Macquarie University, Sydney, New South Wales, Australia

Objective: Determine the relationship between cognitive load (CL) and automation bias (AB).

Background: Clinical decision support (CDS) for electronic prescribing can improve safety but introduces the risk of AB, where reliance on CDS replaces vigilance in information seeking and processing. We hypothesized high CL generated by high task complexity would increase AB errors.

Method: One hundred twenty medical students prescribed medicines for clinical scenarios using a simulated e-prescribing system in a randomized controlled experiment. Quality of CDS (correct, incorrect, and no CDS) and task complexity (low and high) were varied. CL, omission errors (failure to detect prescribing errors), and commission errors (acceptance of false positive alerts) were measured.

Results: Increasing complexity from low to high significantly increased CL, $F(1, 118) = 71.6$, $p < .001$. CDS reduced CL in high-complexity conditions compared to no CDS, $F(2, 117) = 4.72$, $p = .015$. Participants who made omission errors in incorrect and no CDS conditions exhibited lower CL compared to those who did not, $F(1, 636.49) = 3.79$, $p = .023$.

Conclusion: Results challenge the notion that AB is triggered by increasing task complexity and associated increases in CL. Omission errors were associated with lower CL, suggesting errors may stem from an insufficient allocation of cognitive resources.

Application: This is the first research to examine the relationship between CL and AB. Findings suggest designers and users of CDS systems need to be aware of the risks of AB. Interventions that increase user vigilance and engagement may be beneficial and deserve further investigation.

Keywords: automation bias, cognitive load, task complexity, human-computer interaction, working memory, health information technology, medication management and safety, patient safety, medication alerts, compliance, reliance

Address correspondence to David Lyell, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, NSW 2109, Australia; e-mail: david.lyell@mq.edu.au.

HUMAN FACTORS
Vol. 60, No. 7, November 2018, pp. 1008–1021
DOI: 10.1177/0018720818781224

BACKGROUND

One cause of errors when using clinical decision support (CDS) in electronic prescribing (e-prescribing) is automation bias (AB; Lyell et al., 2017). AB has been defined as "the tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing" (Mosier & Skitka, 1996, p. 205). Heuristic replacement is the process in which people use CDS as a mental shortcut rather than using their own cognitive resources. AB can cause clinicians to commit omission errors (a failure to take appropriate action to avoid prescribing errors because they were not alerted to the error by CDS) and commission errors (when clinicians comply with incorrect CDS recommendations; Mosier & Skitka, 1996; Mosier, Skitka, Heers, & Burdick, 1998). Important to note, omission and commission errors are used to classify errors that relate to the task assisted by automation; for CDS this is the detection of prescribing errors.

E-prescribing systems are widely used (Sweidan et al., 2011) and typically include CDS features to reduce the risk that patients will be harmed by prescribing errors (Britt et al., 2015) such as drug-drug interactions (Ammenwerth, Schnell-Inderst, Machan, & Siebert, 2008; van Rosse et al., 2009; Wolfstadt et al., 2008). CDS typically increases the overall quality and safety of prescriptions (Ammenwerth et al., 2008), but like all clinical information technology, it can introduce new classes of error (Kim, Coiera, & Magrabi, 2017), including AB. There are several potential causes of AB (Goddard, Roudsari, & Wyatt, 2012; Parasuraman & Manzey, 2010), such as the presence of highly reliable automation (Bagheri & Jamieson, 2004b; Bailey & Scerbo, 2007; Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006; Molloy & Parasuraman, 1992; Parasuraman, Molloy, & Singh, 1993; Prinzel, Freeman, & Prinzel, 2005; Rovira, McGarry, & Parasuraman, 2007; I. L. Singh, Sharma, &

Parasuraman, 2001; I. L. Singh, Singh, & Saha, 2007; A. L. Singh, Tiwari, & Singh, 2009). This could be due to greater reliability engendering greater trust in automation. Users with high trust in automation are less likely to detect automation failures (Bagheri & Jamieson, 2004a; Bailey & Scerbo, 2007). Task complexity and task difficulty also have been cited as potential factors; however, the evidence is mixed (Bailey & Scerbo, 2007; Goddard et al., 2012; Lyell et al., 2017; Povyakalo, Alberdi, Strigini, & Ayton, 2013). It has been suggested that higher complexity tasks result in AB in a single-task environment, whereas lower complexity tasks result in AB only when performed in a multitask environment (Lyell & Coiera, 2017). For this study, CDS represents a class of automation whose function is to determine whether a prescribing error is present in the prescriptions entered into an e-prescribing system. In contrast, medication alerts are the method by which CDS communicates the discovery of a potential prescribing error to the user. Of particular interest is whether *task complexity* could lead to increased dependence on CDS to manage additional workload and prevent cognitive overload as complexity increases (Sintchenko & Coiera, 2003). Alternatively, AB may be unrelated to task complexity. For example, humans might preferentially offload decision-making effort onto CDS rather than use their own cognitive resources (Mosier & Skitka, 1996).

The present work focuses on testing this complexity hypothesis, using cognitive load theory as the theoretical framework (Lyell & Coiera, 2017). Cognitive load theory is based on the notion that human cognitive ability is limited by the capacity of working memory (Sweller, Ayres, & Kalyuga, 2011), which has limited storage (Cowan, 2001; Miller, 1956) and a short duration (Peterson & Peterson, 1959). When information processing takes place in working memory (Baddeley, 1992), it generates *cognitive load* (Sweller et al., 2011). Cognitive load theorists differentiate between the *intrinsic load* generated by the task and *extraneous load*, which is generated by sources not associated with the goals of the task (Sweller et al., 2011). When cognitive load exceeds available working memory, it can lead to error and hinder learning (Ayres, 2001; Ayres & Sweller, 1990). This study focuses on intrinsic cognitive load, which is expected to be influenced by task complexity. The user interface, which could generate extraneous load, was held constant in the experiment.

This work focuses on the analysis of the relationship between cognitive load and AB. We hypothesize that as task complexity increases, so too will participants' cognitive load, which, in turn, would lead to a greater need to rely on CDS to manage workload. This would be demonstrated by (a) individuals reporting a greater reduction in cognitive load when assisted by CDS than when unaided as complexity increases; (b) increasing omission and commission errors with increased task complexity (Furthermore, it is expected that errors will increase by a greater amount for conditions assisted by CDS that is incorrect compared to unassisted conditions, indicating greater reliance on CDS [although we have previously reported that task complexity may have no effect on AB errors; Lyell et al., 2017]); and (c) errors in unaided conditions being associated with higher cognitive load, which would indicate an association between errors and cognitive overload. Alternatively, if task complexity has no association with these AB-induced errors or errors have no association with higher cognitive load, then this is evidence that AB may be unrelated to task complexity.

METHOD

Experiment Design

This study reports an analysis of data collected as part of a previously reported e-prescribing experiment (Lyell et al., 2017). That study found significantly more participants made omission and commission errors when they prescribed with CDS that was incorrect compared to when they prescribed with no CDS. The present study extends the first by presenting an analysis of how cognitive load was affected by task complexity, CDS, and AB errors. The experiment had two within-subject factors: quality of CDS (correct, incorrect, and no CDS) and task complexity (low, low with interruption, and high), providing a total of nine conditions. Each participant completed one scenario in all nine conditions. The experimental

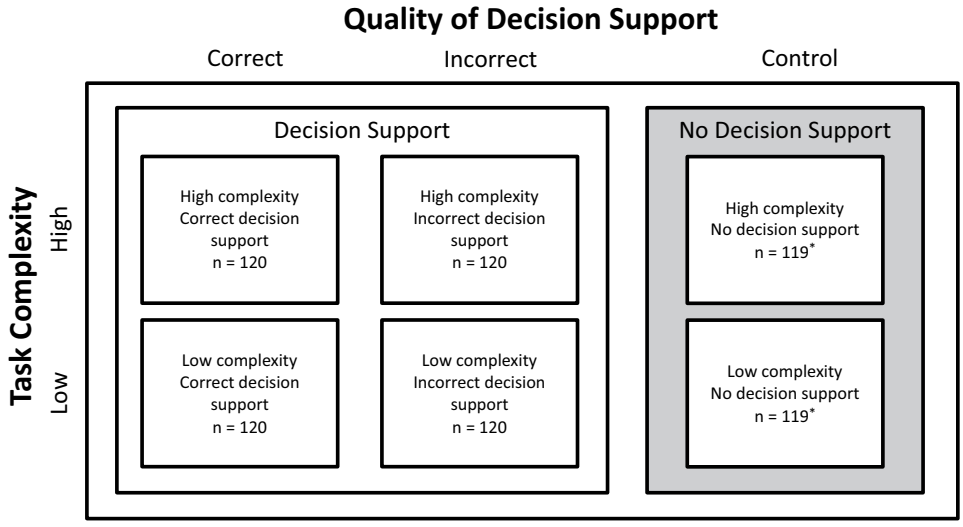

Figure 1. Experimental design with the number of participants in each condition. All participants performed one patient scenario in each condition. The allocation of patient scenarios to conditions was counterbalanced. The order of presentation and whether control conditions were presented first or last were randomized.
*Cognitive load data were not recorded for 1 participant in two control conditions.

control comprised conditions with no CDS; for those conditions, participants were told CDS had been switched off. For the present analysis, we will focus on the low and high task complexity conditions as these aim to manipulate participants' cognitive load (see Figure 1).

Participants

A total of 120 students enrolled in the final 2 years of a medical degree at Australian universities participated in the study. The participants' average age was 24.5 years (*SD* = 2.99), and 53.3% were male. They responded to advertisements emailed by medical schools or posts on social media via medical students' societies. This research complied with the National Statement on Ethical Conduct in Human Research 2007 (National Health and Medical Research Council, Australian Research Council, & Australian Vice-Chancellors' Committee, 2007/2015) and was conducted in accordance with protocols approved by the human research ethics committees of Macquarie University and the University of New South Wales. Informed consent was obtained from each participant, and participants were fully debriefed upon completion of the experiment. Participants were offered two movie vouchers and a certificate for their participation.

One participant completed the experiment twice (on two separate occasions), and the data from the second attempt were excluded. Cognitive load scores were not recorded for two control scenarios affecting 1 participant; the affected scenarios were also excluded from the analysis.

Experimental Task

Participants were tasked with prescribing between three and eight medicines per patient scenario, using an e-prescribing system (see Figure 2). The prescribing system included a decision support function that should generate a warning alert when a prescribing error was detected. Each scenario provided a brief patient history together with a list of the medications participants were asked to prescribe for the hypothetical patient (see Appendix A for examples). Participants were instructed not to prescribe medications that they considered unsafe. One of the requested medications was a deliberately inserted prescribing error that was clearly contraindicated and posed a sufficiently severe risk of harm that it should be

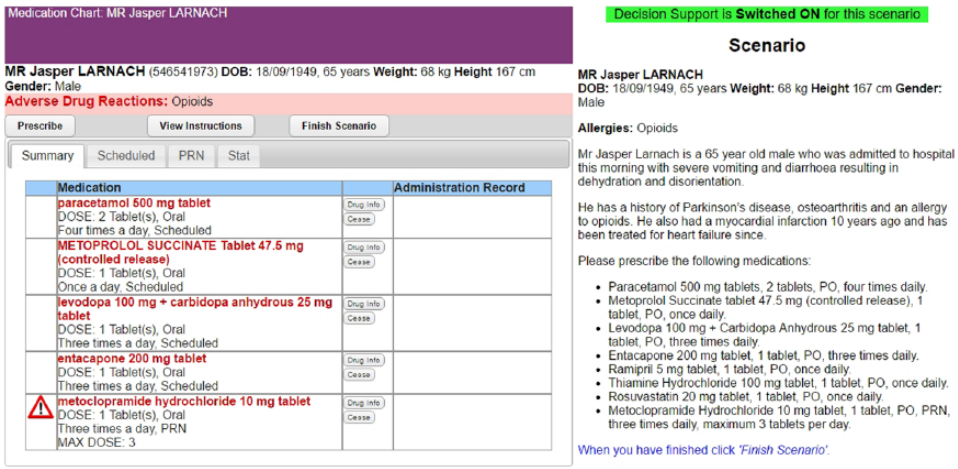

Figure 2. Example of the experimental task showing the e-prescribing system (left) and patient scenario (right).
Source. Lyell et al. (2017). Reproduced under CC BY 4.0.

avoided in all circumstances (see Appendix B). Participants' prescribing this medicine had failed to detect the prescribing error and take appropriate action so was classified as an *omission error.* All other medicines requested in each scenario were safe and appropriate for the patient and therefore should be prescribed. In some of the experimental conditions, an incorrect computer alert was triggered during prescribing, suggesting a safe medication was in error. Participants' acting on an incorrect CDS alert by removing or not prescribing the safe medicine was classified as a *commission error* (wrong action).

The simulated e-prescribing system allowed for the triggering of CDS alerts for prescribing errors, and these alerts were manipulated as follows:

Correct CDS.

- Correct CDS alerts were triggered by the prescription of the unsafe medicine (a true positive). Due to the severity of prescribing errors in the experimental scenarios, all correct alerts were highly relevant.
- The correct absence of an alert when there was no prescribing error indicated a true negative.

Incorrect CDS.

- The incorrect absence of an alert when an unsafe medicine was prescribed (a false negative) provided an opportunity for an omission error if participants failed to detect the error.
- An alert incorrectly warning about a safe medicine (a false positive) provided an opportunity for a commission error.

No CDS (control).

- No CDS served as the control condition in which there was no CDS checking for errors. Participants were informed that decision support had been switched off and was unavailable. They were advised to use instead the inbuilt drug reference to check for any errors they suspected.

A CDS alert was interruptive, displayed as a modal window over the top of the e-prescribing system (see Figure 3). It required a resolution, either by removing the prescription or overriding the alert with a reason, before allowing the participant to proceed. The alert also provided direct access to the relevant drug reference.

Important to note, for errors to be evidence of AB, participants had to be able to perform the task independently of CDS and be able to assess its correctness. The e-prescribing system facilitated this by providing access to a drug information reference (Australian Medicines Handbook, 2015) in all conditions. The reference contained all information necessary to avoid omission and commission errors independently of CDS.

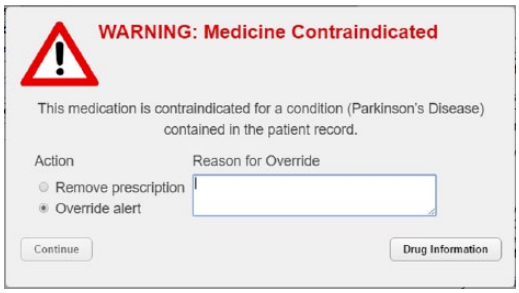

Figure 3. Clinical decision support alert.
Source. Lyell et al. (2017). Reproduced under CC BY 4.0.

Task complexity was manipulated by increasing the number of information elements in each scenario. Information elements were medical conditions, symptoms, test results, prior treatments, allergies, observations, or requested prescriptions that needed to be considered in prescribing decisions. Low-complexity scenarios contained 6 elements, and high-complexity scenarios contained 17. The more elements, the more potential interactions there were between elements that needed to be processed, which should lead to greater complexity and cognitive load (Sweller, 1994).

The allocation of the patient scenarios to experimental conditions was counterbalanced to control for potential differences between scenarios. The order of presentation and whether participants received control scenarios first or last were randomized. The position of errors in the list of medicines to be prescribed was also randomized, allocated at the time of scenario design.

Cognitive Load Measurement

It has been shown that people can reliably self-rate their invested mental effort (Gopher & Braune, 1984), that is, the cognitive resources allocated to a task (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Self-rating of cognitive load (Paas, 1992; Sweller et al., 2011) has been shown to be reliable, unobtrusive, and sensitive to small differences (Paas et al., 2003) and is widely used (Sweller et al., 2011). A cognitive load inventory (see Appendix C) was adapted from a validated instrument (Leppink, Paas, van der Vleuten, van Gog, & van Merriënboer, 2013; Leppink, Paas, van Gog, van der Vleuten, & van Merriënboer, 2014) to reflect the nature of the specific tasks in the present study. The cognitive load inventory was administered at the end of each condition.

The three items measuring intrinsic cognitive load had excellent internal consistency (Cronbach's alpha = .915). These were converted to the same scale and averaged, creating one measure of intrinsic cognitive load for the analysis.

Procedure

The experiment was presented to participants as an evaluation of an e-prescribing system still under development. They were cautioned, "Initial testing has shown that alerts are highly accurate, but occasionally have been incorrect. Therefore, you should always double check with the inbuilt drug information reference." To be consistent with prior AB studies, these instructions emphasized two important points: (a) CDS could be incorrect and (b) the method available to verify CDS uses a nonautomated information source (in this study the drug references; e.g., Bahner, Huper, & Manzey, 2008; Skitka, Mosier, & Burdick, 2000). No information was provided about which types of errors the system would check and alert. Nor were participants provided any information about the prescribing errors inserted into the scenarios. Participants were blinded to the experimental manipulations and experimentally presented opportunities for errors. A 3-min instructional video about how to use the e-prescribing system was shown; it included a demonstration of a correct CDS alert and how to view drug information references.

Participants were instructed as follows: (a) Approach tasks as if they were treating a real patient, exercising all due care. (b) Should they detect any prescribing errors, these should be addressed by not prescribing that medicine. (c) If the error involved an adverse drug interaction between two medicines, only one should be omitted. (d) If there was a discrepancy between CDS and the drug information, they should always rely on the drug information. Participants were not provided any feedback on their performance.

Outcome Measures

Each condition provided the opportunity for participants to make one omission and one

commission error. For each condition, we measured the following:

1. Intrinsic cognitive load (0 to 10).
2. Omission error (yes/no): Participants were scored as making an omission error if they prescribed the medication containing the prescribing error. This indicated they had failed to detect the error and take appropriate action to avoid it. If the participant corrected the error, for example, by reducing a harmful dose to a safe level, it was not scored as an omission error.
3. Commission error (yes/no): Participants were scored as making a commission error if they removed or did not prescribe a safe medicine that received a false positive alert from incorrect CDS (wrong action). As a control, there was a comparator medicine in the correct CDS and control conditions that was safe and received no CDS alert; not prescribing it was scored as a commission error.

Statistical Analyses

Intrinsic cognitive load was measured as a continuous variable, it was normally distributed, and there were no outliers. The effect of task complexity and the presence of decision support on intrinsic cognitive load were tested using a 2 (Task Complexity) × 3 (Quality of Decision Support) repeated-measures analysis of variance. Multivariate results are reported (Wilks's Lambda), which do not assume sphericity.

Differences in cognitive load between participants who did and did not make omission and commission errors were tested in the control condition with independent samples *T* tests. The sample was likely to have low power to detect differences in the CDS-assisted conditions due to the uneven split of observations with and without errors. Therefore, in the event of a significant effect in the control conditions, a multilevel analysis was to be undertaken to leverage more power from the sample. Multilevel models (MLMs) explain sources of variance at multiple levels of analysis (Hoffman & Rovine, 2007). The first level describes differences between whether participants did or did not make an error, and the second level describes the relationship between the experimental factors: task complexity and quality of decision support.

The fixed effects that were assessed for inclusion in MLM were task complexity (low or high), quality of decision support (correct, incorrect, or no CDS), omission error (yes or no), and commission error (yes or no). We also tested all two-way interactions except the interaction between the two error types. A stepwise backward elimination method was used for predictor selection, where all potential predictors were entered into the model and then interactions were removed one by one in order of least significance. The process was repeated for main effects. Model fit was evaluated by comparing models using the likelihood ratio test (Peugh, 2010). Predictors with a significant effect on model fit were retained, while predictors with a nonsignificant effect were discarded. The model included a random intercept for each participant, taking into account the nested structure of the data. All models were estimated using maximum likelihood.

Wherever post hoc pairwise comparisons are reported, the probabilities have been adjusted for multiple comparisons using the Bonferroni correction (Bland & Altman, 1995).

RESULTS

All 120 participants completed one scenario in each of the six experimental conditions. The median time taken for low-complexity scenarios was 2:45 min (interquartile range = 1:42 to 4:08) and for high-complexity scenarios, it was 5:25 min (interquartile range = 3:59 to 7:21).

Intrinsic Cognitive Load Increases With Task Complexity

Increasing task complexity from low to high significantly increased intrinsic cognitive load (see Table 1) with a very large effect size, Wilks's Lambda = .622, $F(1, 118) = 71.597$, $p < .001$, $\eta_p^2 = .378$. In comparison, the quality of decision support, Wilks's Lambda = .925, $F(2, 117) = 4.721$, $p = .011$, $\eta_p^2 = .075$, and the interaction between task complexity and quality of decision support, Wilks's Lambda = .931, $F(2, 117) = 4.367$, $p = .015$, $\eta_p^2 = .069$, also significantly affected intrinsic load but produced a medium effect size (Richardson, 2011).

Incorrect CDS generated a higher cognitive load than correct CDS ($p = .031$). No CDS

TABLE 1: Mean Intrinsic Cognitive Load (Standard Error) by Task Complexity and Quality of Decision Support

	Correct CDS	Incorrect CDS	Control No CDS	Overall Mean
Low complexity	3.4 (.16)	3.8 (.16)	3.6 (.18)	3.6[a] (.13)
High complexity	4.2[b] (.17)	4.4 (.16)	4.9[b] (.19)	4.5[a] (.14)
Overall mean	3.8[c] (.15)	4.1[c] (.14)	4.2 (.16)	4.1 (.13)

Note. Means sharing a common superscript are significantly different from each other ($p < .05$) as determined by a two-way repeated-measures analysis of variance. CDS = clinical decision support.
[a]$p < .001$.
[b]$p = .007$ Bonferroni-corrected pairwise comparison.
[c]p = .031 Bonferroni-corrected pairwise comparison.

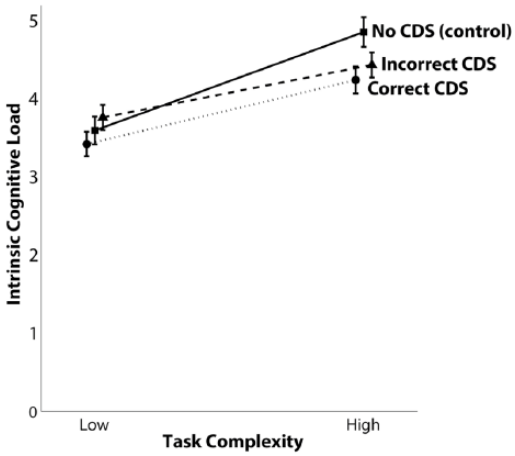

Figure 4. Mean and standard error for intrinsic cognitive load by task complexity and quality of decision support. This shows the interaction between task complexity and quality of decision support.

(control) also generated a higher cognitive load than correct CDS; however, this was not significant ($p = .052$). There was no difference between incorrect CDS and no CDS (control; $p = 1.000$).

The significant interaction between task complexity and quality of decision support (see Figure 4) indicates that the effect of CDS on cognitive load changed depending on task complexity. Correct CDS significantly reduced cognitive load compared to no CDS (control) in high-complexity scenarios ($p = .007$). Cognitive load was lower in high-complexity scenarios with incorrect CDS compared to control conditions; however, this was not significant once the Bonferroni correction was applied ($p = .066$). In low-complexity scenarios, there were no significant differences between the three CDS conditions.

Cognitive Load Differences Between Participants Who Did and Did Not Make Omission and Commission Errors

Participants who made omission errors in control conditions, with no CDS, reported a significantly lower cognitive load than participants who detected the errors (see Table 2) for both low-, $t(117) = 4.087$, $p < .001$, and high-complexity conditions, $t(117) = 2.104$, $p = .038$. This supported performing an MLM analysis. However, no statistically significant differences were found between participants who made and avoided commission errors for low-, $t(117) = .303$, $p = .701$, and high-complexity conditions, $t(117) = .202$, $p = .840$.

Multilevel Analysis: Omission Errors Were Associated With Lower Cognitive Load

We evaluated 11 models; from these five fixed effects, we found them to significantly contribute to the fit of MLM (see Appendix D) and included them in the analysis (see Table 3 for the significance of fixed effects and Appendix E for model coefficients). The final model was significantly better than the intercepts-only model, $\chi^2(8) = 110.431$, $p < .001$.

TABLE 2: Mean Intrinsic Cognitive Load and Standard Deviation by Whether Participants Made Omission or Commission Errors in Each Experimental Condition

		No Error		Error	
Clinical Decision Support Condition	Task Complexity	*n*	Cognitive Load Mean (*SD*)	*n*	Cognitive Load Mean (*SD*)
Failure to detect prescribing errors (omission errors)					
Correct	Low	111	3.44 (1.7)	9	3.32 (2.0)
Incorrect	Low	25	4.19 (1.8)	95	3.66 (1.7)
Control	Low	65	4.22 (1.7)	54	2.85 (1.9)
Correct	High	109	4.25 (1.9)	11	4.31 (1.9)
Incorrect	High	30	4.97 (1.6)	90	4.26 (1.8)
Control	High	62	5.24 (2.0)	57	4.44 (2.1)
False positive errors (commission errors)					
Correct	Low	115	3.39 (1.7)	5	4.33 (1.0)
Incorrect	Low	37	4.00 (2.0)	83	3.66 (1.6)
Control	Low	115	3.61 (1.9)	4	3.23 (2.6)
Correct	High	108	4.30 (1.8)	12	3.88 (2.1)
Incorrect	High	43	4.97 (1.6)	77	4.14 (1.8)
Control	High	104	4.87 (2.1)	15	4.76 (1.9)

TABLE 3: Significance of Fixed Effects in the Multilevel Model of Intrinsic Cognitive Load

	df	*F*	*p*
Intercept	1, 163.216	983.781	<.001*
Task complexity (low complexity, high complexity)	1, 597.611	82.500	<.001*
Quality of decision support (correct CDS, incorrect CDS, control [no CDS])	2, 621.989	0.971	.379
Omission error (omission error, no omission error)	1, 667.989	4.230	.040*
Complexity × Decision Support	1, 597.666	3.584	.028*
Decision Support × Omission Error	1, 636.490	3.785	.023*

Note. CDS = clinical decision support.
*$p < .05$.

Commission errors did not significantly contribute to model fit and were therefore excluded from the analysis. The intraclass correlation coefficient was .406, indicating that 40.6% of the variance in intrinsic load was attributable to variation between participants, supporting the conduct of a multilevel analysis (Hayes, 2006; Twisk, 2006). The model residuals were normally distributed.

Overall, participants who correctly avoided making omission errors reported a significantly higher cognitive load ($M = 4.3$, $SE = .14$) than those who made omission errors ($M = 4.0$, $SE = .17$). There also was a significant interaction between omission errors and the quality of decision support (see Figure 5). Participants who avoided omission errors reported significantly higher cognitive load than participants who made errors with incorrect CDS ($p = .012$) and no CDS (control; $p < .001$).

DISCUSSION

To better understand the cause of AB, we conducted a randomized controlled experiment that

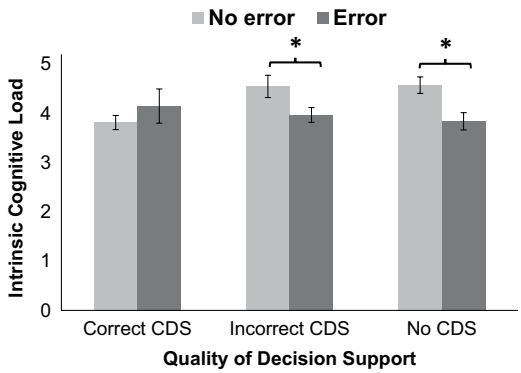

Figure 5. Intrinsic cognitive load, mean and standard error by omission error, and quality of decision support. This illustrates the interaction between quality of decision support and whether the participant made an omission error. Asterisks indicate significant differences ($p < .05$), determined by the multilevel model.

focused on the relationship between task complexity and AB. We varied task complexity and the correctness of decision support, measuring omission and commission errors, as well as intrinsic cognitive load. In doing so, we sought to determine whether AB results from increased reliance on decision support due to rising task complexity mediated by cognitive load. For this hypothesis to be supported, we would have needed to find evidence of (a) more AB errors in high- than in low-complexity conditions, (b) errors in control scenarios associated with higher cognitive load, and (c) a greater reduction in cognitive load in CDS-assisted compared to -unassisted conditions as task complexity increases. However, the present findings do not support this hypothesis. While there was a greater reduction in cognitive load between CDS-assisted and -unassisted conditions in high- compared to low-complexity scenarios, we have previously reported this experiment found no difference in error rates between low and high complexity, in any decision support condition (Lyell et al., 2017). Moreover, omission errors were associated with lower, not higher, intrinsic cognitive load in both incorrect CDS and no CDS conditions (see Figure 5).

The finding that omission errors are associated with lower cognitive load is at odds with cognitive load theory, which typically associates errors with higher cognitive load (Ayres, 2001; Ayres & Sweller, 1990), where the demands of the task either exceed cognitive capacity or do not leave sufficient capacity to enable learning. However, recalling that our measurements of cognitive load are based on self-rated mental effort, errors may have occurred because the cognitive resources participants allocated to the task fell short of those required by it.

AB and Cognitive Load

This experiment found a significant relationship between omission errors and cognitive load. Taken together, the findings—that (a) incorrect CDS induced more omission errors than no CDS and (b) lower cognitive load is associated with omission errors across both conditions—seem to suggest that when people suffer an AB, there is a reduction in the cognitive load allocated to the task.

The greatest reductions in cognitive load were observed in CDS-assisted, compared to -unassisted conditions, but only when task complexity was high (see Figure 4). This could represent participants' use of CDS to reduce the complexity and cognitive demands of high-complexity tasks (Sintchenko & Coiera, 2003). CDS did not provide the same benefit when task complexity was low. While this suggests that increasing complexity might pose an increased risk of AB, it had no impact on observed AB errors.

Prior AB studies provide support for the idea that automation is associated with a reduction in cognitive resources allocated to a task. Workload can be reduced by higher levels of automation (Manzey, Reichenbach, & Onnasch, 2012; Rovira et al., 2007) and more reliable automation (Prinzel et al., 2005; Wickens, Clegg, Vieane, & Sebok, 2015). Metzger and Parasuraman (2005) and Reichenbach, Onnasch, and Manzey (2011) found an improvement in secondary task performance, suggesting that automation may enable allocation of resources to secondary tasks without affecting overall workload (Metzger & Parasuraman, 2005). These studies measured workload with the NASA-TLX, which incorporates six dimensions (mental, physical, temporal, performance, frustration,

and effort; Hart & Staveland, 1988). In contrast, intrinsic cognitive load measures mental effort applied to processing the complexity of the task, a likely subcomponent of workload but a useful metric for assessing the effects of task complexity. We are not aware of any reporting on differences in workload between individuals who do and do not make AB errors within the same experimental conditions. In the present study, this analysis was made possible by the measurement of cognitive load per CDS condition. This contrasts with other AB studies, which typically measured workload per experimental block, recording a rate of errors, and where participants received mostly correct but some incorrect automation (e.g., Prinzel et al., 2005; Reichenbach et al., 2011).

There was no association between cognitive load and commission errors, suggesting that omission and commission errors may involve different cognitive processes and demands. Indeed, Bahner, Elepfandt, and Manzey (2008) have suggested that omission and commission errors are separate and independent phenomena, having observed that people were differently affected by decision support false negatives and false positives.

Reduced cognitive load appears to be one way in which AB manifests in relation to omission errors. This supports the cognitive miser hypothesis of AB (Mosier & Skitka, 1996), which assumes that humans have limited cognitive capacity. It describes the preference for people to seek adequate, faster, and less effortful ways of thinking, rather than engaging in more accurate but slower and more effortful thinking (Fiske & Taylor, 1984), that is, to travel the path of least cognitive effort. This is achieved through the use of mental shortcuts or heuristics. This is consistent with Mosier and Skitka's (1996) definition of AB as the use of automation as a heuristic, relying on it instead of engaging in information seeking and processing. Research exists showing AB is associated with a reduction in information seeking (Bagheri & Jamieson, 2004a; Bahner, Elepfandt, et al., 2008; Bahner, Huper, et al., 2008; Manzey et al., 2012). We believe this study provides the first evidence of an association between AB and reduced intrinsic cognitive load. Our findings, however, relate only to omission errors.

Implications

We hypothesized that users of automated decision aids would make more AB errors as their tasks became more complex. We also hypothesized that the cause of these increased errors was the increased cognitive load we assumed would be generated by increasingly complex tasks (Lyell & Coiera, 2017).

Surprisingly, our experimental results were not consistent with these hypotheses. Intriguingly, cognitive load appeared to have an inverse relationship with omission errors. Lower, rather than higher, cognitive load was associated with omission errors. There was no statistically significant association between cognitive load and commission errors.

Just as surprisingly, while we found the expected association between increased complexity and increased cognitive load, complexity itself was not associated with changes in omission errors. This suggests either that complexity and load are relatively independent of each other in their association with omission errors or that the results require replication to ensure there were no issues with statistical sampling or study design.

If these results hold, then individuals assign cognitive resources independently of the complexity of the task they are undertaking. This implies that some tasks are assigned more resources than needed—a safe but inefficient situation—but also that some tasks receive fewer resources than needed, leading to errors. This may explain why lower cognitive loads were associated with omission errors.

Pragmatically, there are several implications of these results. First, they reconfirm the challenges of using automated decision aids and the caveat that both designers and users of such systems need to be aware of the risks of AB errors. However, interventions tested to date, including providing training on AB and how to avoid errors (Mosier, Skitka, Dunbar, & McDonnell, 2001), exposure to examples of automation failures in training (Bahner, Huper, et al., 2008), and externally imposed accountability for performance (Mosier et al., 1998; Skitka et al., 2000), have had limited success in reducing AB errors.

Next, the results show low cognitive load levels are most likely to trigger omission errors.

Interventions that assist individuals to recognize the amount of cognitive effort they need to invest in a particular task may be beneficial. For example, interventions that increase user engagement during periods of low cognitive load or that increase vigilance and engagement deserve investigation.

Finally, the causal relationships between task complexity, cognitive load, and AB errors are more complicated than initially assumed. In particular, the triggers for high- or low-cognitive-load states seem to extend beyond the specific needs of the task and appear to demonstrate more complex strategies for the allocation of cognitive resources.

Limitations

The design of this experiment was subject to several limitations. First, the inclusion of opportunities for both omission and commission errors in the same condition was necessary due to the limited number of trials that could be presented without participants dropping out of the experiment. However, it limited our ability to fully differentiate between the cognitive load effects arising from omission and commission errors. More research is needed to determine the effect of false positive alerts on cognitive load.

Second, the use of medical students provided a necessary control, ensuring that participants had an equivalent level of knowledge and expertise. These are factors that affect how participants are affected by task complexity, which in turn affects their cognitive load. These results are limited in generalizability; they are likely to be representative of new clinicians entering practice but not of experienced clinicians.

Third, this study used a single task, meaning that participants' attention was not divided between multiple, concurrently performed tasks. AB is also commonly reported in multiple tasks (Lyell & Coiera, 2017), wherein the allocation of attention between tasks is an important factor (Parasuraman & Manzey, 2010). More research is needed to determine the impact of AB on cognitive load when attention is divided between multiple concurrent tasks.

Fourth, this was a controlled experiment, meaning that participants were not subject to time or other external pressures that clinicians would ordinarily experience while prescribing. This may have affected decisions about reliance on CDS.

Finally, the failure of the task complexity manipulation to alter the rate of AB errors limited our ability to assess the relationship between task complexity, cognitive load, and AB. There were omission and commission errors in both high- and low-complexity tasks. If task complexity truly is a cause of AB, then it is likely the low-complexity condition in this experiment exceeded the threshold at which AB presents. Future research could further enhance our understanding by replicating this study using a lower level of task complexity at which AB errors are likely to be reduced or eliminated. This would allow an assessment of how cognitive load changes as complexity varies between levels where AB does and does not present.

CONCLUSION

This study sought to understand the relationship between task complexity and AB, using cognitive load theory as a framework. We did not find any evidence of errors resulting from high cognitive load brought about by increased task complexity, which would have indicated that errors resulted from the cognitive demands of the task overwhelming available cognitive resources. Instead, we found that omission errors with incorrect CDS were associated with reduced intrinsic cognitive load. In addition, participants who made omission errors with no CDS exhibited the same reduction in cognitive load but to a significantly lesser extent. This could suggest that omission errors stem from an insufficient allocation of cognitive resources to the task.

Compared to unaided conditions, the extent to which CDS allowed participants to reduce their cognitive effort was greatest in high-complexity conditions. This suggests that increased complexity poses an increased risk of omission errors. However, despite this, there was no effect of complexity on errors. This indicates that task complexity alone is an insufficient cue to trigger the appropriate allocation of cognitive resources and therefore, in isolation, is likely to be a poor predictor of AB.

ACKNOWLEDGMENTS

This research was supported by a doctoral scholarship for David Lyell provided by the HCF Research Foundation, Sydney, Australia. We acknowledge the contributions of Magdalena Z. Raban, L. G. Pont, Melissa T. Baysari, Richard O. Day, John Sweller, Ouhao Chen, Vitaliy Kim, Jingbo Liu, Peter Petocz, Thierry Wendling, Robin Butterfield, Monish Maharaj, and Rhonda Siu as well as the medical students who participated in this study.

David Lyell conceived this research and designed and conducted the study with guidance and under the supervision of Enrico Coiera and Farah Magrabi.

David Lyell drafted the manuscript with input from all authors. All authors provided revisions for intellectual content. All authors have approved the final manuscript.

This study received ethical approval from Macquarie University Human Research Ethics Committee (5201401029) and the University of New South Wales Human Research Ethics Advisory Panel (2014-7-32).

KEY POINTS

- High task complexity significantly increased cognitive load but had no effect on AB errors. This challenges the notion that AB is triggered by increasing task complexity.
- Omission errors were associated with a significant reduction in cognitive load, suggesting that these errors may arise from insufficient allocation of cognitive resources.
- Interventions that might assist users to reduce omission errors are worthy of consideration. Such interventions may need to (a) assist users in recognizing the requisite amount of cognitive resources needed for a task and (b) increase user vigilance.

ORCID ID

David Lyell https://orcid.org/0000-0002-2695-0368

SUPPLEMENTAL MATERIAL

The online supplementary material is available with the manuscript on the *HF* Web site.

REFERENCES

Ammenwerth, E., Schnell-Inderst, P., Machan, C., & Siebert, U. (2008). The effect of electronic prescribing on medication errors and adverse drug events: A systematic review. *Journal of the American Medical Informatics Association*, *15*(5), 585–600. doi:10.1197/jamia.M2667

Australian Medicines Handbook Pty Ltd. (2015, January). *Australian medicines handbook 2015*. Retrieved from http://amhonline.amh.net.au/

Ayres, P., & Sweller, J. (1990). Locus of difficulty in multistage mathematics problems. *American Journal of Psychology*, *103*(2), 167–193. doi:10.2307/1423141

Ayres, P. L. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology*, *26*(2), 227–248. doi:10.1006/ceps.2000.1051

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Bagheri, N., & Jamieson, G. A. (2004a). *Considering subjective trust and monitoring behavior in assessing automation-induced "complacency."* Mahwah, NJ: Lawrence Erlbaum.

Bagheri, N., & Jamieson, G. A. (2004b, October 10–13). *The impact of context-related reliability on automation failure detection and scanning behaviour*. Paper presented at the IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583), The Hague, the Netherlands.

Bahner, J., Elepfandt, M. F., & Manzey, D. (2008, September 22–26). *Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias*. Paper presented at the Human Factors and Ergonomics Society annual meeting, New York, NY.

Bahner, J., Huper, A.-D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, *66*(9), 688–699. doi:10.1016/j.ijhcs.2008.06.001

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, *8*(4), 321–348. doi:10.1080/14639220500535301

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors*, *48*, 693–709. doi:10.1518/001872006779166280

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ*, *310*(6973), 170. doi:10.1136/bmj.310.6973.170

Britt, H., Miller, G., Henderson, J., Bayram, C., Harrison, C., Valenti, L., . . .Charles, J. (2015). *General practice activity in Australia 2014–15. General practice series no. 38*. Sydney, Australia: Sydney University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral Brain Science*, *24*, 87–185.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. Reading, MA: Addison-Wesley.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, *19*(1), 121–127. doi:10.1136/amiajnl-2011-000089

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, *26*, 519–532. doi:10.1177/001872088402600504

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research*, *32*(4), 385–410.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–117. doi:10.3758/bf03192848

Kim, M. O., Coiera, E., & Magrabi, F. (2017). Problems with health information technology and their effects on care delivery and patient outcomes: A systematic review. *Journal of the American Medical Informatics Association*, *24*(2), 246–250. doi:10.1093/jamia/ocw154

Leppink, J., Paas, F., van der Vleuten, C. P., van Gog, T., & van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072.

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P., & van Merriënboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, *30*, 32–42.

Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: A systematic review. *Journal of the American Medical Informatics Association*, *24*(2), 423–431. doi:10.1093/jamia/ocw105

Lyell, D., Magrabi, F., Raban, M. Z., Pont, L. G., Baysari, M. T., Day, R. O., & Coiera, E. (2017). Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making*, *17*(1), 1–10. doi:10.1186/s12911-017-0425-5

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*, *6*(1), 57–87. doi:10.1177/1555343411433844

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors*, *47*, 35–49. doi:10.1518/0018720053653802

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Molloy, R., & Parasuraman, R. (1992, October 12–16). *Monitoring automation failures: Effects of automation reliability and task complexity*. Paper presented at the Human Factors and Ergonomics Society annual meeting, Atlanta, GA.

Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 201–220). Hillsdale, NJ: Lawrence Erlbaum.

Mosier, K. L., Skitka, L. J., Dunbar, M., & McDonnell, L. (2001). Aircrews and automation bias: The advantages of teamwork? *International Journal of Aviation Psychology*, *11*(1), 1–14. doi:10.1207/s15327108ijap1101_1

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology*, *8*(1), 47–63. doi:10.1207/s15327108ijap0801_3

National Health and Medical Research Council, Australian Research Council, & Australian Vice-Chancellors' Committee. (2015). *National statement on ethical conduct in human research*. Retrieved from www.nhmrc.gov.au/guidelines/publications/e72 (Original work published 2007)

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. doi:10.1037/0022-0663.84.4.429

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. doi:10.1207/S15326985EP3801_8

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*, 381–410. doi:10.1177/0018720810376055

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency." *International Journal of Aviation Psychology*, *3*(1), 1–23. doi:10.1207/s15327108ijap0301_1

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. doi:10.1037/h0049234

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, *48*(1), 85–112. doi:10.1016/j.jsp.2009.09.002

Povyakalo, A. A., Alberdi, E., Strigini, L., & Ayton, P. (2013). How to discriminate between computer-aided and computer-hindered decisions: A case study in mammography. *Medical Decision Making*, *33*(1), 98–107. doi:10.1177/0272989x12465490

Prinzel, L. J., III, Freeman, F. G., & Prinzel, H. D. (2005). Individual differences in complacency and monitoring for automation failures. *Individual Differences Research*, *3*(1), 27–49.

Reichenbach, J., Onnasch, L., & Manzey, D. (2011). Human performance consequences of automated decision aids in states of sleep loss. *Human Factors*, *53*, 717–728. doi:10.1177/0018720811418222

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. doi:10.1016/j.edurev.2010.12.001

Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, *49*, 76–87. doi:10.1518/001872007779598082

Singh, A. L., Tiwari, T., & Singh, I. L. (2009). Effects of automation reliability and training on automation-induced complacency and perceived mental workload. *Journal of the Indian Academy of Applied Psychology*, *35*, 9–22.

Singh, I. L., Sharma, H. O., & Parasuraman, R. (2001). Effects of manual training and automation reliability on automation induced complacency in flight simulation task. *Psychological Studies, 46*(1/2), 21–27.

Singh, I. L., Singh, A. L., & Saha, P. K. (2007, July 22–27). *Monitoring performance and mental workload in an automated system*. Paper presented at the International Conference on Engineering Psychology and Cognitive Ergonomics, Beijing, China.

Sintchenko, V., & Coiera, E. W. (2003). Which clinical decisions benefit from automation? A task complexity approach. *International Journal of Medical Informatics*, *70*(2/3), 309–316. doi:10.1016/S1386-5056(03)00040-6

Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human Computer Studies*, *52*(4), 701–717. doi:10.1006/ijhc.1999.0349

Sweidan, M., Williamson, M., Reeve, J. F., Harvey, K., O'Neill, J. A., Schattner, P., & Snowdon, T. (2011). Evaluation of features to support safety and quality in general practice clinical software. *BMC Medical Informatics and Decision Making*, *11*(1), 1–8. doi:10.1186/1472-6947-11-27

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. doi:10.1016/0959-4752(94)90003-5

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory* (Vol. 1). New York, NY: Springer.

Twisk, J. (2006). *Applied multilevel analysis: A practical guide*. Cambridge, UK: Cambridge University Press.

van Rosse, F., Maat, B., Rademaker, C. M. A., van Vught, A. J., Egberts, A. C. G., & Bollen, C. W. (2009). The effect of computerized physician order entry on medication prescription errors and clinical outcome in pediatric and intensive care: A systematic review. *Pediatrics*, *123*(4), 1184–1190. doi:10.1542/peds.2008-1494

Wickens, C. D., Clegg, B. A., Vieane, A. Z., & Sebok, A. L. (2015). Complacency and automation bias in the use of imperfect automation. *Human Factors*, *57*, 728–739. doi:10.1177/0018720815581940

Wolfstadt, J. I., Gurwitz, J. H., Field, T. S., Lee, M., Kalkar, S., Wu, W., & Rochon, P. A. (2008). The effect of computerized physician order entry with clinical decision support on the rates of adverse drug events: A systematic review. *Journal of General Internal Medicine*, *23*(4), 451–458. doi:10.1007/s11606-008-0504-5

David Lyell is a PhD candidate at the Centre for Health Informatics at Macquarie University. He received his MBA in 2006 from Deakin University.

Farah Magrabi is an associate professor of health informatics at the Centre for Health Informatics at Macquarie University. She received her PhD in biomedical engineering in 2003 from the University of New South Wales.

Enrico Coiera is a professor of health informatics and the director of the Centre for Health Informatics at Macquarie University. He received his PhD in computer science in 1990 from the University of New South Wales.

Date received: August 27, 2017
Date accepted: May 14, 2018

Article III: Appendices

APPENDIX A: EXAMPLES OF PATIENT SCENARIOS

Background

The following excerpts are taken from Lyell, D., Magrabi, F., Raban, M. Z., Pont, L. G., Baysari, M. T., Day, R. O., & Coiera, E. (2017). Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making, 17*(1), 28. doi:10.1186/s12911-017-0425-5.

Australian medical education uses an integrative approach where students learn patient and clinical content throughout their degree. By the final two years of their education, participants would have typically received training in rational and safe prescribing. They also complete the National Prescribing Curriculum, a series of online modules based on the prescribing principles outlined in the World Health Organisation's Guide to Good Prescribing (De Vries, Henning, Hogerzeil, & Fresle, 1994). Upon completion of these final two years, graduates would begin practice as junior medical officers.

The prescribing scenarios were developed with advice from an expert panel, including four hospital doctors, a medical pharmacology registrar and two pharmacists (including MZR). They were independently reviewed by a consultant physician specialising in pharmacology (RD), to ensure clinical relevance. The scenarios presented hypothetical patient scenarios and involved prescribing tasks that were typical of those undertaken by junior medical officers, based on observations of e-prescribing in a medical ward of a major teaching hospital. A common task performed by junior medical officers is the prescribing of medications using an e-prescribing system upon admission of a patient to hospital, including medicines taken prior to, and those initiated on admission.

Each scenario included one genuine prescribing error, where one of the medicines was clinically contraindicated in that scenario (Appendix B). These were designed to be unambiguously errors and of sufficient severity in the risk posed to the patient that the medicine should be avoided under all circumstances. To ensure this, the severity of the errors included in the scenarios were independently assessed by a clinical pharmacist (LGP). The error in one scenario was assessed as potentially lethal, five were serious, and three were significant (Dornan et al., 2009). All other medicines listed in scenarios where carefully chosen so as to be unambiguously free from error.

Low complexity scenario (Scenario D)

MR Thomas Chapman
DOB: 21/05/1971, 43 years **Weight:** 103 kg **Height** 176 cm **Gender:** Male
Allergies: Nil

Mr Thomas Chapman is a 43 year old man who presented in the emergency department with a severe headache and vomiting that have persisted for the last 24 hours.

He suffers from peptic ulcer disease with a history of bleeds.

Please prescribe the following medications:

- Pantoprazole 40 mg tablet: enteric, 1 tablet, PO, once daily.
- Aspirin 300 mg tablet: effervescent, 3 tablets, PO, every 6 hours.
- Metoclopramide Hydrochloride 10 mg tablet, 1 tablet, PO, PRN, every 8 hours, maximum 3 tablets per day.

Box 1 The text of patient scenario D shown to participants.

Prescribing error: Aspirin is contraindicated for this patient who suffers peptic ulcer disease as it increases the risk of gastrointestinal ulceration (Australian Medicines Handbook Pty Ltd, 2015 January). The severity of this error was assessed as significant. Aspirin should not be prescribed to this patient.

Omission error: Participants who prescribed Aspirin (which was contra-indicated) failed to detect the error and take appropriate action to avoid it. If Aspirin was prescribed in this scenario it was scored as an omission error.

Commission error: Participants made a commission error if they accepted a false alert from incorrect CDS, warning that Pantoprazole was contra-indicated in patients with severe vomiting, by stopping or not prescribing that medicine (wrong action). The alert was factually incorrect and contradicted by the drug reference. Pantoprazole is a medicine used to treat peptic ulcer disease (Australian Medicines Handbook Pty Ltd, 2015 January). It is safe, appropriate and should be prescribed to the patient in this scenario who suffers peptic ulcer disease. If Pantoprazole was not prescribed in this scenario it was scored as a commission error. As a control measure, this scoring applied to all conditions, regardless of whether a false alert was displayed.

Medicine	Correct response	Error	Correct CDS	Incorrect CDS	Control
Pantoprazole	Safe and appropriate Should be prescribed	Commission error if not prescribed	No alert *true-negative*	Alert *false-positive*	No alert No CDS support
Aspirin	Contraindicated Should not be prescribed	Omission error if prescribed	Alert *true-positive*	No alert *false-negative*	No alert No CDS support
Metoclopramide Hydrochloride	Safe and appropriate Should be prescribed	-	No alert *true-negative*	No alert *true-negative*	No alert No CDS support

Table 1 Medicines requested for the patient in Scenario D.

High complexity scenario (Scenario I)

MR Jasper Larnach
DOB: 18/09/1949, 65 years **Weight:** 68 kg **Height** 167 cm **Gender:** Male
Allergies: Opioids

Mr Jasper Larnach is a 65 year old male who was admitted to hospital this morning with severe vomiting and diarrhoea resulting in dehydration and disorientation.

He has a history of Parkinson's disease, osteoarthritis and an allergy to opioids. He also had a myocardial infarction 10 years ago and has been treated for heart failure since.

Please prescribe the following medications:

- Paracetamol 500 mg tablets, 2 tablets, PO, four times daily.
- Metoprolol Succinate tablet 47.5 mg (controlled release), 1 tablet, PO, once daily.
- Levodopa 100 mg + Carbidopa Anhydrous 25 mg tablet, 1 tablet, PO, three times daily.
- Entacapone 200 mg tablet, 1 tablet, PO, three times daily.
- Ramipril 5 mg tablet, 1 tablet, PO, once daily.
- Thiamine Hydrochloride 100 mg tablet, 1 tablet, PO, once daily.
- Rosuvastatin 20 mg tablet, 1 tablet, PO, once daily.
- Metoclopramide Hydrochloride 10 mg tablet, 1 tablet, PO, PRN, three times daily, maximum 3 tablets per day.

Box 2 The text of patient scenario I shown to participants.

Prescribing error: Metoclopramide Hydrochloride is contraindicated for this patient who suffers Parkinson's disease as it may cause their symptoms to worsen (Australian Medicines Handbook Pty Ltd, 2015 January). The severity of this error was assessed as serious. Metoclopramide Hydrochloride should not be prescribed to this patient.

Omission error: Participants who prescribed Metoclopramide Hydrochloride (which was contra-indicated) failed to detect the error and take appropriate action to avoid it. If Metoclopramide Hydrochloride was prescribed in this scenario it was scored as an omission error.

Commission error: Participants made a commission error if they accepted a false alert in the Incorrect CDS condition, warning of an adverse drug interaction between Entacapone and Rosuvastatin, by stopping or not prescribing one or both of those medicines (wrong action). The alert was factually incorrect and contradicted by the drug reference. Entacapone is a medicine used in the treatment of Parkinson's disease, and Rosuvastatin is a medicine used to treat hypercholesterolaemia (high blood cholesterol; Australian Medicines Handbook Pty Ltd, 2015 January) a risk for people with coronary heart disease. Both medicines are safe, appropriate and should be prescribed for the patient in this scenario who suffers Parkinson's disease and is being treated for heart failure. If Entacapone or Rosuvastatin were not prescribed in this scenario it was scored as a commission error. As a control measure, this scoring applied to all conditions, regardless of whether a false alert was displayed.

Medicine	Correct response	Error	Correct CDS	Incorrect CDS	Control
Paracetamol	Safe and appropriate Should be prescribed	-	No alert *true-negative*	Alert *false-positive*	No alert No CDS support
Metoprolol Succinate	Safe and appropriate Should be prescribed	-	No alert *true-negative*	No alert *true-negative*	No alert No CDS support
Levodopa	Safe and appropriate Should be prescribed	-	No alert *true-negative*	No alert *true-negative*	No alert No CDS support
Entacapone	Safe and appropriate Should be prescribed	Commission error if <u>not</u> prescribed	No alert *true-negative*	Alert when prescribed with Rosuvastatin *false-positive*	No alert No CDS support
Ramipril	Safe and appropriate Should be prescribed	-	No alert *true-negative*	No alert *true-negative*	No alert No CDS support
Thiamine Hydrochloride	Safe and appropriate Should be prescribed	-	No alert *true-negative*	No alert *true-negative*	No alert No CDS support
Rosuvastatin	Safe and appropriate Should be prescribed	Commission error if <u>not</u> prescribed	No alert *true-negative*	Alert when prescribed with Entacapone false-positive	No alert No CDS support
Metoclopramide Hydrochloride	Contraindicated Should <u>not</u> be prescribed	Omission error if prescribed	Alert *true-positive*	No alert *false-negative*	No alert No CDS support

Table 2 Medicines requested for the patient in Scenario I.

Please note: *Patients presented in the prescribing scenarios are fictional. All biographical information was made up for the purpose of this experiment in order to present participants with the information they would expect in such patient cases.*

APPENDIX B: OVERVIEW OF ERRORS IN PATIENT SCENARIOS

Scenario	Error Type	Medication	Alert displayed to participants	Comment	Severity
A **Low complexity**	Prescribing Error	Digoxin 250 microgram tablet, 2 tablets, PO, three times a day.	WARNING: High Dose The entered dose is higher than the recommended maintenance dose range.	The elderly patient has atrial fibrillation which was controlled with Digoxin prior to admission. The dose requested by the scenario is a loading dose. The maintenance dose for an elderly patient is 62.5 to 125 micrograms once daily.	Serious
	False Positive	Lisinopril 5mg tablet, 1 tablet, PO, once daily.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Pulmonary Oedema) contained in the patient record.	Lisinopril is not contraindicated in patients with pulmonary oedema.	
B **Low complexity**	Prescribing Error	Spironolactone 25mg tablet, 1 tablet, PO, once daily.	WARNING: Medicine Contraindicated This medicine is contraindicated for a condition (Hyperkalaemia) contained in the patient record.	Patient has hyperkalemia for which Spironolactone is contraindicated.	Serious
	False Positive	Augmentin Duo Forte (amoxycillin 875 mg + clavulanic acid 125 mg) tablet, 1 tablet, PO, twice daily.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Systolic Heart Failure) contained in the patient record.	Augmentin Duo Forte is not contraindicated in patients with heart failure.	
C **Low complexity**	Prescribing Error	Warfarin Sodium 2 mg tablet, 1 tablet, PO, once daily. and Ibuprofen 400 mg tablet, 1 tablet, PO, three times daily.	WARNING: Adverse Drug Interaction This medication has a listed adverse interaction with another already prescribed medication. There is an adverse drug interaction for Warfarin and Ibuprofen. This combination should be avoided.	Non-steroidal anti-inflammatory drugs (Ibuprofen) increase the risk of gastrointestinal bleeding in patients taking Warfarin. The combination should be avoided, especially as better analgesic options are available.	Significant
	False Positive	Atorvastatin 10 mg tablet, 1 tablet, PO, once daily.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Venous Thromboembolism) contained in the patient record.	Atorvastatin is not contraindicated in patients with venous thromboembolism.	
D **Low complexity**	Prescribing Error	Aspirin 300 mg tablet: effervescent, 3 tablets, PO, every 6 hours.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Peptic Ulcer Disease) contained in the patient record.	Patient has peptic ulcer disease with a history of bleeds for which aspirin increases the risk of gastrointestinal ulceration. There are better analgesic options.	Significant
	False Positive	Pantoprazole 40 mg tablet: enteric, 1 tablet, PO, once daily.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Severe Vomiting) contained in the patient record.	Pantoprazole is not contraindicated in patients with server vomiting.	
E **Low complexity**	Prescribing Error	Loperamide Hydrochloride 2 mg capsule, 1 capsule, PO, PRN, every four hours, maximum 8 capsules per day.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Ulcerative Colitis) contained in the patient record.	Loperamide is contraindicated in patients with ulcerative colitis which poses a risk of toxic megacolon.	Serious
	False Positive	Mesalazine 500 mg tablet: enteric, 1 tablet, PO, three times daily. and Prednisolone 25 mg tablet, 1 tablet, PO, once daily.	WARNING: Adverse Drug Interaction This medication has a listed adverse interaction with another already prescribed medication. There is an adverse drug interaction for Mesalazine and Prednisolone. This combination should be avoided.	There is no documented adverse drug interaction for Mesalazine and Prednisolone.	

Scenario	Error Type	Medication	Alert displayed to participants	Comment	Severity
F **Low complexity**	Prescribing Error	Phenelzine 15 mg tablet, 1 tablet, PO, three times daily. and Tramadol Hydrochloride 50mg capsules, 2 capsules, PO, PRN, every six hours, maximum 8 capsules per day.	WARNING: Adverse Drug Interaction This medication has a listed adverse interaction with another already prescribed medication. There is an adverse drug interaction for Phenelzine and Tramadol hydrochloride. This combination should be avoided.	The combination of phenelzine and tramadol are contraindicated due to the possibility of causing serotonin toxicity.	Serious
	False Positive	Ramipril 5 mg tablet, 1 tablet, PO, once daily.	WARNING: Adverse Drug Reaction This patient has an Allergy or Adverse Drug Reaction recorded for this medication.	The patient is allergic to Sulfonamide. However Ramipril is not contraindicated for this allergy.	
G **High complexity**	Prescribing Error	Paracetamol 500 mg tablet, 2 tablets, PO, four times a day. and Panadeine Forte (Codeine Phosphate with Paracetamol Tablet 30 mg-500 mg) tablet, 2 tablets, PO, PRN, every four hours, maximum 8 tablets per day.	WARNING: High Dose / Duplicate Substance Both Paracetamol and Panadine Forte (Codeinie Phosphate 30mg with Paracetamol 500mg) contain the ingredient Paracetamol. The total Paracetamol entered is higher than the recommended dose range.	Prescribed together these two prescriptions provide for a combined maximum possible dose of 8 grams of paracetamol per day, double the maximum daily dose of 4 grams.	Significant
	False Positive	Ciprofloxacin 250 mg tablet, 1 tablet, PO, twice daily.	WARNING: Adverse Drug Reaction This patient has an Allergy or Adverse Drug Reaction recorded for this medication.	The patient is allergic to penicillin. Ciprofloxacin is an antibiotic however it is not contraindicated for allergy to penicillin.	
H **High complexity**	Prescribing Error	Methotrexate 2.5 mg tablets, 3 tablets, PO, once daily.	WARNING: High Dose The entered dose is higher than the recommended maintenance dose range	Patient has net onset rheumatoid arthritis. For treatment of rheumatoid arthritis, the loading dose of methotrexate is 7.5mg once weekly.	Potentially lethal
	False Positive	Paracetamol 500 mg tablet, 2 tablets, PO, PRN, every four hours, maximum 8 tablets per day.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Peptic Ulcer Disease) contained in the patient record.	Patient has newly diagnosed peptic ulcer disease, however it is not a contraindication for paracetamol.	
I **High complexity**	Prescribing Error	Metoclopramide Hydrochloride 10 mg tablet, 1 tablet, PO, PRN, three times daily, maximum 3 tablets per day.	WARNING: Medicine Contraindicated This medication is contraindicated for a condition (Parkinson's Disease) contained in the patient record.	Patient has a history of Parkinson's disease for which Metoclopramide is contraindicated as symptoms may worsen. The drug reference(2015 January) names an alternative medicine as being preferred.	Serious
	False Positive	Entacapone 200 mg tablet, 1 tablet, PO, three times daily. and Rosuvastatin 20 mg tablet, 1 tablet, PO, once daily.	WARNING: Adverse Drug Interaction This medication has a listed adverse interaction with another already prescribed medication. There is an adverse drug interaction for Entacapone and Rosuvastatin. This combination should be avoided.	There is no documented adverse drug interaction for Entacapone and Rosuvastatin.	

APPENDIX C: COGNITIVE LOAD INVENTORY

Item IL1 – expected to measure intrinsic cognitive load

Item: The content of this scenario was very complex.
Original: The content of this activity was very complex.

Scale: 0 (Not at all the case), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (Completely the case)

Item IL2 – expected to measure intrinsic cognitive load

Item: I invested a very high mental effort in the complexity of this scenario.
Original: I invested a very high mental effort in the complexity of this activity.

Scale: 0 (Not at all the case), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (Completely the case)

Item IL3 – expected to measure intrinsic cognitive load

Item: The scenario I just finished was...
Original: The lecture that just finished was…

Scale: Very, very easy (1), Very easy (2), Easy (3), Rather easy (4), Neither easy nor difficult (5), Rather difficult (6), Difficult (7), Very difficult (8), Very, very difficult (9)

APPENDIX D: SELECTION OF PREDICTORS FOR MULTILEVEL MODEL

Model Evaluated	Fixed effect removed	-2 Log Likelihood	Number of Parameters	Likelihood ratio test	Include in final model
Fixed Intercept, Random Intercept (Participant), Complexity, CDS, Omission Error, Commission Error, Complexity*CDS, Complexity*Omission Error, Complexity*Commission Error, CDS*Omission Error, CDS*Commission Error		2673.228	16		
Fixed Intercept, Random Intercept (Participant), Complexity, CDS, Omission Error, Commission Error, Complexity * CDS, Complexity * Commission Error, CDS * Omission Error, CDS * Commission Error	Complexity*Omission Error	2673.243	15	$\chi^2(1)=0.015$, p=.903	Exclude
Fixed Intercept, Random Intercept (Participant), Complexity, CDS, Omission Error, Commission Error, Complexity*CDS, CDS*Omission Error, CDS*Commission Error	Complexity*Commission Error	2673.505	14	$\chi^2(1)=0.262$, p=.609	Exclude
Fixed Intercept, Random Intercept (Participant), Complexity, CDS, Omission Error, Commission Error, Complexity*CDS, CDS*Omission Error	CDS*Commission Error	2674.962	12	$\chi^2(2)=1.457$, p=.483	Exclude
Fixed Intercept, Random Intercept (Participant), Complexity, CDS, Omission Error, Commission Error, CDS*Omission Error	Complexity*CDS	2682.599	10	$\chi^2(2)=7.637$, p=.022	Include
Fixed Intercept, Random Intercept (Participant), Complexity, CDS, Omission Error, Commission Error	CDS*Omission Error	2690.058	8	$\chi^2(2)=7.459$, p=.024	Include
Fixed Intercept, Random Intercept (Participant), Complexity, CDS, Omission Error	Commission Error	2690.785	7	$\chi^2(1)=0.727$, p=.394	Exclude
Fixed Intercept, Random Intercept (Participant), Complexity, CDS	Omission Error	2703.03	6	$\chi^2(1)=12.245$, p<.001	Include
Fixed Intercept, Random Intercept (Participant), Complexity	CDS	2712.908	4	$\chi^2(1)=9.878$, p=.007	Include
[Intercepts only model] Fixed Intercept, Random Intercept (Participant)	Complexity	2786.906	3	$\chi^2(1)=73.998$, p<.001	Include
Fixed Intercept	Random Intercept	2966.725	2	$\chi^2(1)=179.819$, p<.001	Include

Final model

Intercept for fixed effects, Random Intercept: Participant (Covariance structure: Variance Components), Task Complexity (Low, High), Quality of Clinical Decision Support (Correct, Incorrect, No CDS), Omission error (Yes, No), Quality of Clinical Decision Support * Omission Error, Scenario Complexity * Quality of Clinical Decision Support, Residual.

11 parameters, -2 Log likelihood = 2676.475.

APPENDIX E: INTRINSIC COGNITIVE LOAD MULTILEVEL MODEL COEFFICIENTS

	Coefficient	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	4.48	.194	508.517	23.038	<.001	4.10	4.86
Low Complexity	-1.28	.175	597.6	-7.337	<.001	-1.62	-0.94
High Complexity	.	.	.	.	.	.	.
Correct CDS	0.08	.371	632.675	0.211	.833	-0.65	0.81
Incorrect CDS	-0.19	.207	609.211	-0.915	.360	-0.60	0.22
Control (No CDS)	.	.	.	.	.	.	.
No Omission Error	0.73	.194	652.848	3.778	<.001	0.35	1.12
Omission Error	.	.	.	.	.	.	.
Low Complexity * Correct CDS	0.46	.246	597.595	1.883	.060	-0.02	0.95
Low Complexity * Incorrect CDS	0.64	.246	597.705	2.591	.010	0.15	1.12
Low Complexity * Control (No CDS)	.	.	.	.	.	.	.
High Complexity * Correct CDS	.	.	.	.	.	.	.
High Complexity * Incorrect CDS	.	.	.	.	.	.	.
High Complexity * Control (No CDS)	.	.	.	.	.	.	.
Correct CDS * No Omission Error	-1.07	.392	640.594	-2.73	.007	-1.84	-0.30
Correct CDS * Omission Error	.	.	.	.	.	.	.
Incorrect CDS * No Omission Error	-0.16	.290	632.583	-0.534	.593	-0.73	0.41
Incorrect CDS * Omission Error	.	.	.	.	.	.	.
Control (No CDS) * No Omission Error	.	.	.	.	.	.	.
Control (No CDS) * Omission Error	.	.	.	.	.	.	.

. = parameter is redundant.

REFERENCES

Australian Medicines Handbook Pty Ltd. (2015 January). Australian Medicines Handbook 2015 (online). Retrieved from http://amhonline.amh.net.au/

De Vries, T. P. G. M., Henning, R. H., Hogerzeil, H. V., & Fresle, D. A. (1994). *Guide to good prescribing*. Geneva: World Health Organization.

Dornan, T., Ashcroft, D., Heathfield, H., Lewis, P., Miles, J., Taylor, D., . . . Wass, V. (2009). An in-depth investigation into causes of prescribing errors by foundation trainees in relation to their medical education: EQUIP study. *London: General Medical Council*, 1-215.

Chapter 4 summary

4.5 Effect of task complexity and clinical decision support on cognitive load

4.5.1 Omission errors

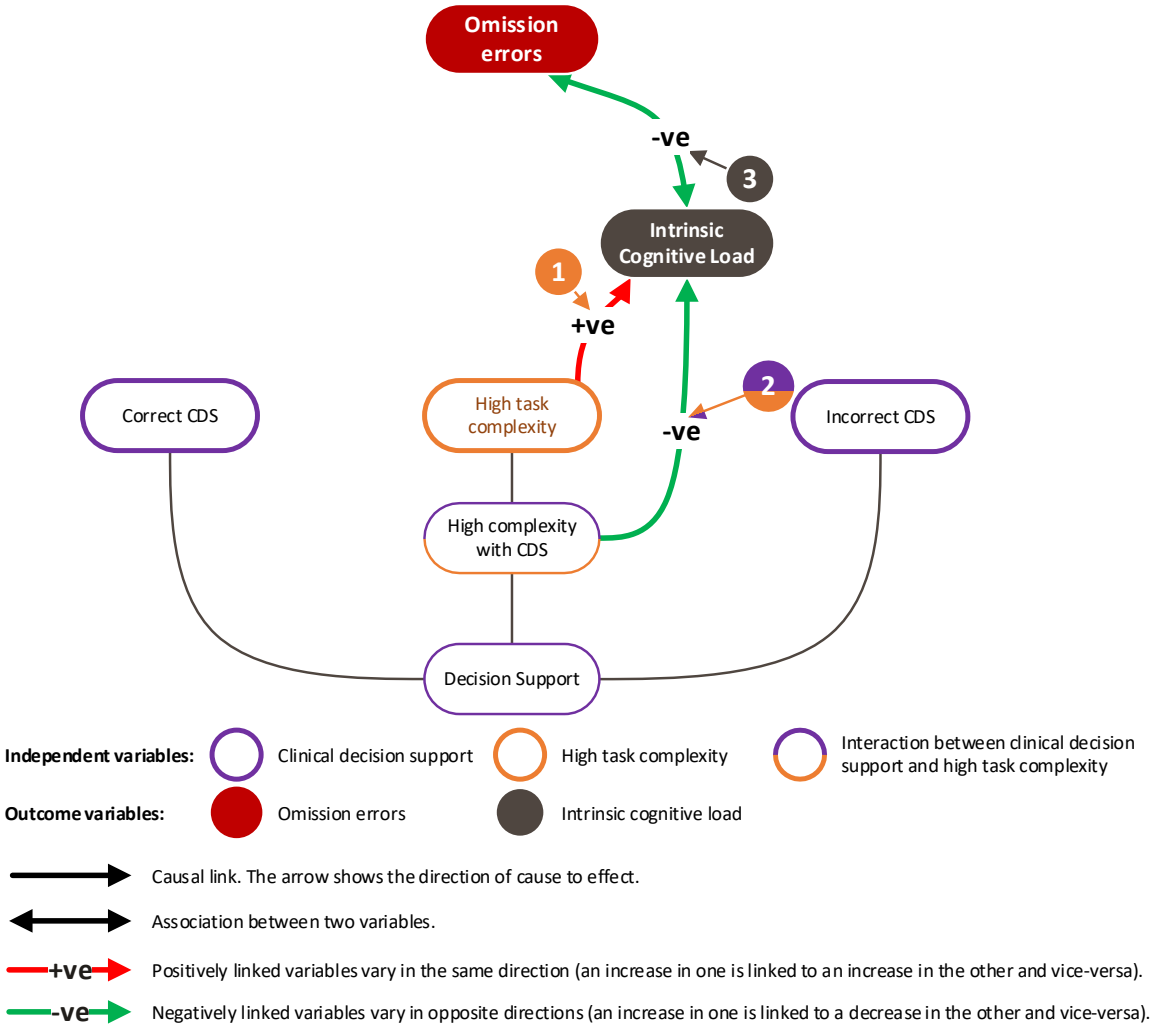

Figure 4-1 Effect of task complexity, clinical decision support and omission errors on intrinsic cognitive load

1 High task complexity increases intrinsic cognitive load.

Intrinsic cognitive load was significantly higher for high- compared to low-complexity scenarios. See Table 1 in Article III.

2 Intrinsic cognitive load was reduced in high complexity conditions assisted by CDS.

The presence of CDS reduced intrinsic cognitive load compared to when there was no CDS in high-, but not in low-complexity scenarios. This was significant for correct CDS, however, became non-significant for incorrect CDS once the Bonferroni correction was applied. See Figure 4 and Table 1 in Article III.

3 Reduced omission errors were associated with increased intrinsic cognitive load

Participants who avoided omission errors with incorrect CDS and no CDS reported significantly higher intrinsic cognitive load than participants who made omission errors. See Figure 5 in Article III.

4.5.2 Commission errors

The analysis revealed no association between reported intrinsic cognitive load and commission errors.

4.6 Chapter 4 References

1. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association* 2017;**24**(2):423-31 doi: 10.1093/jamia/ocw105

2. Lyell D, Magrabi F, Raban MZ, Pont LG, Baysari MT, Day RO, Coiera E. Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making* 2017;**17**(1):28 doi: 10.1186/s12911-017-0425-5

3. Lyell D, Magrabi F, Coiera E. The effect of cognitive load and task complexity on automation bias in electronic prescribing. *Human Factors* 2018;**60**(7):1008 - 21 doi: 10.1177/0018720818781224

4. Li SYW, Magrabi F, Coiera E. A systematic review of the psychological literature on interruption and its patient safety implications. *Journal of the American Medical Informatics Association* 2011;**19**(1):6-12 doi: 10.1136/amiajnl-2010-000024

5 Automation bias and verification

This chapter examines the relationship between automation bias and verification.

5.1 Background

In the e-prescribing experiment, participants could assess the safety and appropriateness of medicines using the provided drug reference.[1] This drug reference provided the means to identify unsafe medicines independent of CDS. This was essential in the control condition where there was no CDS support. It could also be used to verify the correctness of CDS alerts or their absence.

This chapter examines participants' information seeking or verification, as measured by their access of the drug reference, and how it relates to automation bias errors and task complexity.

5.2 Contributions of this article to thesis

Article IV reports on the analysis of participants' access of drug references during the electronic prescribing experiment. It reports that less verification was significantly associated with both omission and commission errors. It also found that verification was reduced by the presence of CDS and high task complexity.

The findings reported in this article (Article IV) contribute to an understanding of the relationship between automation bias errors, task complexity and verification. This contributes to aim 6, which will be addressed in the discussion (chapter 6).

5.3 Article details

This article was published in *Applied Clinical Informatics*.[2]

Citation

Lyell D, Magrabi F, Coiera E. Reduced Verification of Medication Alerts Increases Prescribing Errors. *Applied Clinical Informatics* 2019;**10**(01):066-76 doi: 10.1055/s-0038-1677009

The version of record is available from the publisher's website:

https://doi.org/10.1055/s-0038-1677009

5.4 Author contributions

David Lyell conceived this research and designed and conducted the study with guidance from, and under the supervision of, Enrico Coiera and Farah Magrabi.

David Lyell drafted the manuscript with input from all authors. All authors provided revisions for intellectual content. All authors have approved the final manuscript.

Article IV: Reduced verification of medication alerts increases prescribing errors

Reduced verification of medication alerts increases prescribing errors.

David Lyell

Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, NSW 2109 Australia. Telephone: +61 2 9850 2434 Email: david.lyell@mq.edu.au

Farah Magrabi

Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, NSW 2109 Australia. Email: farah.magrabi@mq.edu.au

Enrico Coiera

Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, NSW 2109 Australia. Email: enrico.coiera@mq.edu.au

Correspondence: David Lyell

Keywords:

Automation bias
Human-computer interaction
Clinical Decision Support Systems
Medication alerts
Cognitive biases
Medication Errors

Word count: 4,070

ABSTRACT

OBJECTIVE: Clinicians using clinical decision support (CDS) to prescribe medications have an obligation to ensure prescriptions are safe. One option is to verify the safety of prescriptions if there is uncertainty, e.g. by using drug references. Supervisory control experiments in aviation and process control have associated errors with reduced verification arising from over-reliance on decision support. However, it is unknown whether this relationship extends to clinical decision making. Therefore, we examine whether there is a relationship between verification behaviours and prescribing errors, with and without CDS medication alerts, and whether task complexity mediates this.

METHOD: One hundred and twenty students in the final two years of a medical degree prescribed medicines for patient scenarios using a simulated e-prescribing system. CDS (correct, incorrect and no CDS) and task complexity (low and high) were varied. Outcomes were omission (missed prescribing errors) and commission errors (accepted false-positive alerts). Verification measures were access of drug references and view time percentage of task time.

RESULTS: Failure to access references for medicines with prescribing errors increased omission errors with no CDS (high-complexity: $\chi^2(1)=12.716$, $p<.001$) and incorrect CDS (Fisher's exact; low-complexity: $p=.002$, high-complexity: $p=.001$). Failure to access references for false-positive alerts increased commission errors (low-complexity: $\chi^2(1)=16.673$, $p<.001$, high-complexity: $\chi^2(1)=18.690$, $p<.001$). Fewer participants accessed relevant references with incorrect compared to no CDS (McNemar; low-complexity: $p<.001$, high-complexity: $p<.001$). Lower view time percentages increased omission, $F(3, 361.914)=4.498$, $p=.035$, and commission errors, $F(1, 346.223)=2.712$, $p=.045$. View time percentages were lower in CDS-assisted compared to unassisted conditions, $F(2, 335.743)=10.443$, $p<.001$.

DISCUSSION: The presence of CDS reduced verification of prescription safety. When CDS was incorrect, reduced verification was associated with increased prescribing errors.

CONCLUSION: CDS can be incorrect, and verification provides one mechanism to detect errors. System designers need to facilitate verification without increasing workload or eliminating the benefits of correct CDS.

BACKGROUND

Prescribing errors are a leading cause of preventable adverse drug events.[1] A common cause of prescribing errors is a lack of knowledge about medicines and the patients for whom they are being prescribed.[2] Clinical decision support (CDS) within electronic prescribing (e-prescribing) systems has been shown to reduce adverse events by alerting clinicians to potential errors such as drug-drug interactions.[3-5] However, CDS is not a perfect substitute for information about medicines: not all potential problems are alerted,[6] malfunctions can occur,[7-9] and alerts are frequently overridden.[10, 11]

Verification is the process of establishing the truth or correctness of something by investigation or evaluation of data.[12] Prescribing errors could be avoided by verification of prescriptions, testing their correctness (safety and appropriateness) against information published in drug references. Inadequate verification is considered an indicator of complacency in overseeing automation, such as decision support.[13-15]

Of specific concern, clinicians may over-rely on CDS and consequently reduce their verification efforts, which could lead to errors when CDS is incorrect. This over-reliance is known as automation bias and occurs when CDS alerts are used as a "*heuristic replacement for vigilant information seeking and processing.*"[16] Omission errors occur when clinicians fail to address problems because they were not alerted to the problem by CDS, and commission errors occur when incorrect CDS advice is acted upon.[16-18] Reduced verification has been associated with automation bias errors in the heavily automated domains of aviation and process control in supervisory control tasks,[13-15, 19-22] but has not yet been tested for CDS medication alerts, where tasks, decision support and task complexity are likely to differ.[23]

The evidence for higher task complexity increasing automation bias errors is mixed.[17, 24-26] However high complexity tasks typically have more information to verify [27] and so might result in increased reliance on CDS.[23]

While verification could have a key role in reducing prescribing errors, this relationship has not yet been directly studied. Accordingly, this study examines: (1) the relationship between verification and prescribing errors with and without CDS medication alerts, and (2) whether task complexity mediates this relationship. We are especially interested in the impact of incorrect CDS, which creates the potential for automation bias errors.

METHOD

This study presents an analysis of verification data collected as part of a previously reported e-prescribing experiment.[17] That earlier study reported significant evidence of automation bias, with overreliance on incorrect CDS resulting in significantly more errors than when there was no CDS. A

second analysis evaluated whether high cognitive load was a cause of automation bias, but instead found participants who made omission errors experienced significantly lower cognitive load than those who didn't make errors.[28] This third study extends the prior studies by examining how the presence of CDS and automation bias impact participants' verification and how those changes might contribute to errors.

Participants

Students enrolled in the final two years of a medical degree at Australian universities, who would typically have received training in rational and safe prescribing, and completed the National Prescribing Curriculum, a series of online modules based on the principles outlined in the World Health Organisation's Guide to Good Prescribing.[29]

Experiment design

The analysis had two within-subjects factors: quality of CDS (correct, incorrect and no CDS) and scenario complexity (low and high). The control involved scenarios with no CDS. The original experiment included an interruption condition which was excluded from this analysis as participants were interrupted while verifying.[17] All participants performed one scenario in each of the six conditions (Figure 1).

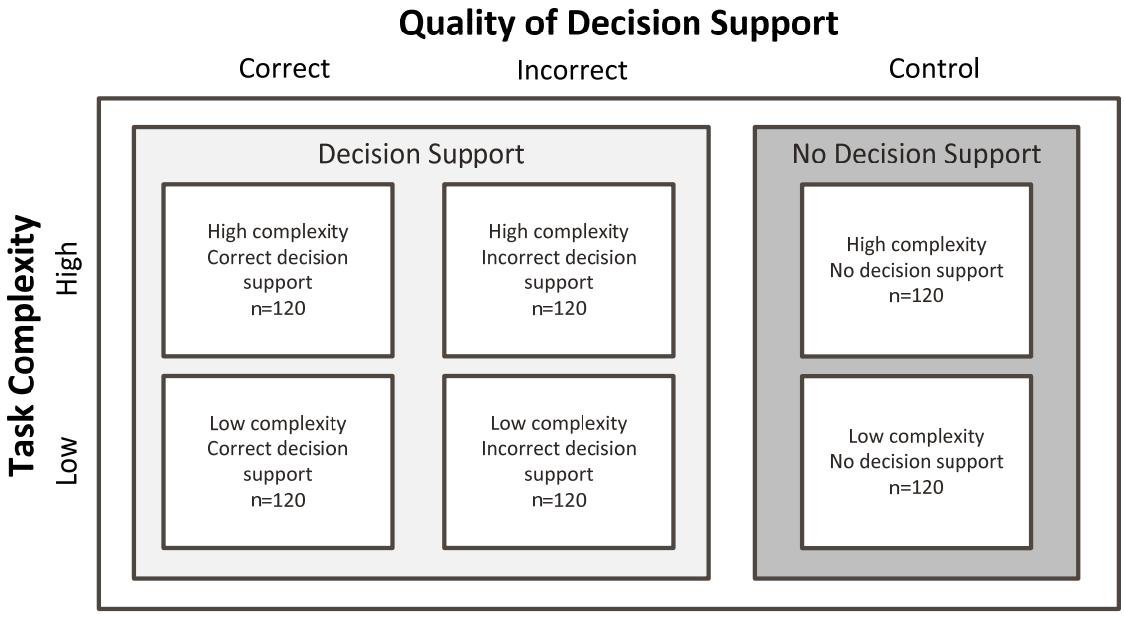

Figure 1 Experimental design with the number of participants in each condition.
(Adapted from "Automation bias in electronic prescribing" by Lyell et al., 2017, BMC Med Inform Decis Mak, 17:28. Adapted and reproduced under CC BY 4.0.)

Outcome measures

1. Omission error (yes/no): Participants made an *omission error* if they prescribed a designated medication containing a prescribing error, indicating that they had failed to detect it. If the error was corrected, it was not scored as an error.

2. Commission error (yes/no): Participants made a *commission error* if they wrongly acted on a false-positive alert by not prescribing a medication that was unaffected by prescribing errors.

Verification measures

3. Access (accessed/not accessed): Whether the participant accessed the drug reference for the medicine with the prescribing error (omission error) or the medicine triggering the false-positive alert (commission error).
4. View time percentage: The percentage of task time viewing drug references. The conversion to a percentage of task time allowed for comparisons between low- and high-complexity conditions, which differed in the number of prescription requested. High-complexity scenarios requested five more prescriptions than low-complexity scenarios. Task and drug reference view time were expected to increase as a function of the number of requested prescriptions.

Experimental task

Participants were provided with patient scenarios presenting a brief patient history and a list of medications for them to prescribe using a simulated e-prescribing system (Figure 2). One of the listed medicines was contraindicated, posing a sufficiently severe risk of harm to the patient that its use should be avoided. All other requested medication orders were unaffected by prescribing errors. Participants were instructed to prescribe all medications except those they believed to contain a prescribing error. Of interest was whether participants would detect the prescribing error. See the appendices of Lyell, et al. [28] for examples of the patient scenarios and a summary of the errors inserted in the scenarios.

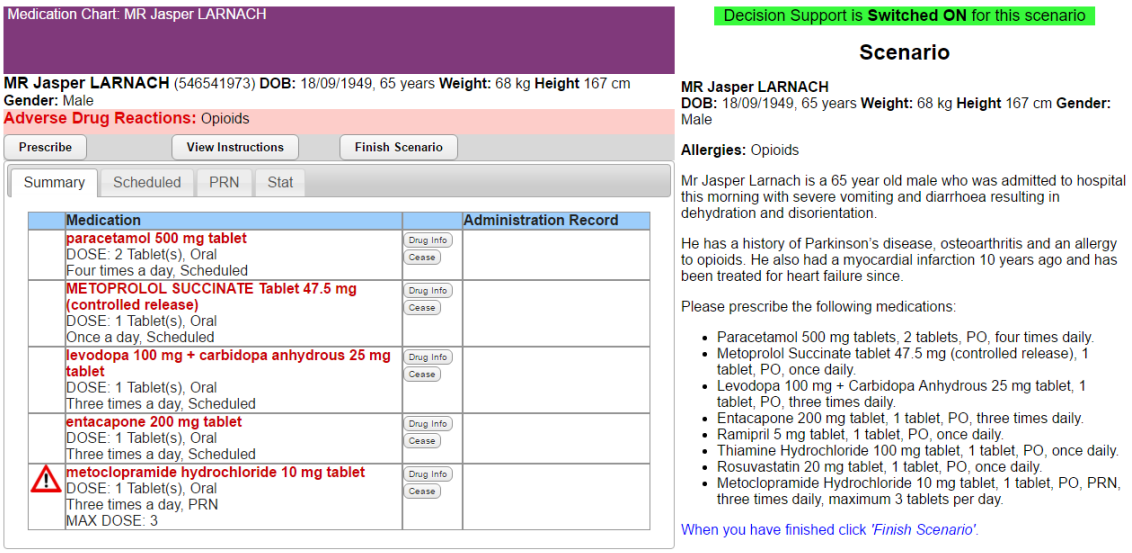

Figure 2 Example of the experimental task showing the e-prescribing system (left) and patient scenario (right).
(From "Automation bias in electronic prescribing" by Lyell et al., 2017, BMC Med Inform Decis Mak, 17:28. Reproduced under CC BY 4.0.)

Verification of prescriptions

Participants were able to verify the safety of prescriptions independently of CDS and the correctness of CDS by accessing a drug reference viewer built into the e-prescribing system. The drug reference was easily accessible and displayed monographs from the Australian Medicines Handbook,[30] an evidence-based reference widely utilised in Australian clinical practice.[31] Participants were instructed: (1) CDS could be incorrect; (2) how to verify using the drug reference; (3) rely on the drug reference over CDS if there was a discrepancy; and (4) refer only to the provided drug reference.

Drug references were checked by MZR (a pharmacist) and DL to ensure they provided clear and sufficient information to enable prescribing errors to be identified. A log recorded access to drug references and view times.

Clinical decision support alerts

CDS displayed alerts (Figure 3) when a medication order containing a prescribing error was entered and required resolution either by removing the prescription or by overriding the alert with a reason. For examples of the override reasons provided by participants see Lyell, et al. [17]

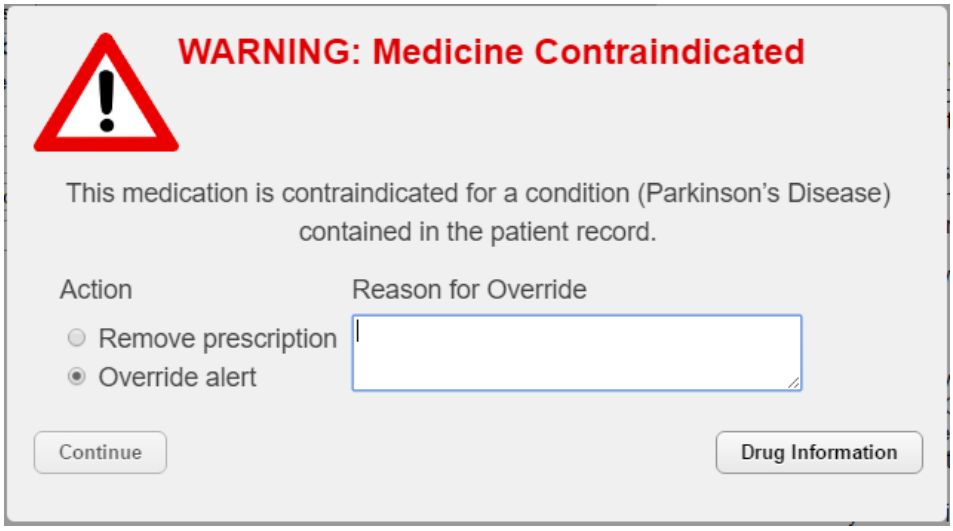

Figure 3 CDS medication alert.
(From "Automation bias in electronic prescribing" by Lyell et al., 2017, BMC Med Inform Decis Mak, 17:28. Reproduced under CC BY 4.0.)

The triggering and content of CDS alerts were manipulated across the three conditions:

- *Correct CDS* alerts were triggered by prescription of the medication with the prescribing error (true-positives). The absence of alerts always indicated true-negatives.
- *Incorrect CDS* failed to alert the prescribing error (false-negative), and instead provided one false-positive alert for a medicine unaffected by prescribing error. These CDS errors provided opportunities for one omission and one commission error.

- *No CDS* served as the control condition in which there were no CDS checking for errors. Participants were told that CDS had been switched off for these scenarios and were advised to use the drug reference to manage any errors.

Task complexity was manipulated by varying the number requested prescriptions and information elements in scenarios.[32, 33] Low-complexity scenarios requested three prescriptions and contained three additional information elements, such as medical conditions, symptoms, test results, allergies and observations that could potentially contra-indicate those medications. High-complexity scenarios requested eight medications and contained nine additional information elements. As a result, high-complexity scenarios had five more drug references which could be viewed, more information elements to be cross-referenced and required more verification than low-complexity scenarios. We previously reported that participants found high-complexity scenarios significantly more cognitively demanding than low-complexity scenarios.[28]

Allocation of patient scenarios to experimental conditions was counterbalanced to ensure scenarios were evenly presented in all conditions. The order of presentation was randomised to control for order effects.

Procedure

The experiment was presented as an evaluation of an e-prescribing system in development. No information was provided on what types of errors the system would check and alert. Participants were shown an instructional video on how to use the e-prescribing system, including demonstration of a correct CDS alert and how to verify using the drug reference.

Participants were instructed to approach tasks as if treating a real patient, exercising all due care and to not prescribe any medication believed to contain a prescribing error.

Statistical analyses

Chi-Square test for independence or Fisher's exact probability tests were used to test whether access of drug references relevant to errors was associated with omission and commission errors. Differences in access between CDS conditions and levels of task complexity were tested with McNemar's tests.

Multilevel modelling,[34] which is not affected by missing data,[35] was used to analyse view time percentage as participants did not access drug references in all conditions. The predictors assessed for inclusion in the model were task complexity, quality of decision support, and whether the participant made an omission error and commission error. We assessed all two-way interactions. A stepwise backward elimination method was used for predictor selection, where all predictors were entered into the model, and then interactions were removed one by one in order of least significance. The process was repeated for main effects. Model fit was evaluated by comparing

models using the likelihood ratio test.[36] Only predictors with a significant effect on model fit were retained. The model included a random intercept for each participant, taking into account the nested structure of the data. Models were constructed using maximum likelihood for parameter estimation.

RESULTS

One hundred and twenty participants were included in the analysis. One participant completed the experiment twice (on two separate occasions), and the data from their second attempt were excluded. Participants' average age was 24 years, and 46.7% were female. The median time to perform low-complexity scenarios was 2:45 min (interquartile range = 1:42 to 4:08) and for high-complexity scenarios, it was 5:25 min (interquartile range = 3:59 to 7:21). Overall participants accessed the drug information reference at least once in 64.7% of scenarios. Thirty-four participants viewed at least one reference in all scenarios, while eleven participants did not view any references (accounting for 25.9% of the scenarios in which no references were viewed).

Accessing drug references for medicines with prescribing errors

Omission errors were higher when drug references for medicines with prescribing errors were not accessed

When prescribing without CDS (control), omission errors were higher when drug references for medicines with prescribing errors were not accessed (Table 1). This was significant for high- (χ^2 (1, n = 120) = 12.716, p <.001, phi = -.326), but not for low-complexity scenarios (χ^2 (1, n = 120) = 1.569, p = .210).

Table 1 Percentage (number) of participants who accessed the drug reference for medicines with prescribing errors by whether an omission error was made.

	Control (No CDS)			Correct CDS			Incorrect CDS		
	No error	Error	Total	No error	Error	Total	No error	Error	Total
Low Complexity									
Accessed	59.7%	40.3%	51.7% (62)	94.1%	5.9%	42.5% (51)	47.6%	52.4%	17.5% (21)
Not accessed	48.3%	51.7%	48.3% (58)	91.3%	8.7%	57.5% (69)	15.2%	84.8%	82.5% (99)
Total	54.2% (65)	45.8% (55)		92.5% (111)	7.5% (9)		20.8% (25)	79.2% (95)	
High Complexity									
Accessed	70.6%	29.4%	42.5% (51)	90.9%	9.1%	45.8% (55)	62.5%	37.5%	13.3% (16)
Not accessed	37.7%	62.3%	57.5% (69)	90.8%	9.2%	54.2% (65)	19.2%	80.8%	86.7% (104)
Total	51.7% (62)	48.3% (58)		90.8% (109)	9.2% (11)		25% (30)	75% (90)	
Total									
Accessed	64.6%	35.4%	47.1% (113)	92.5%	7.5%	44.2% (106)	54.1%	45.9%	15.4% (37)
Not accessed	42.5%	57.5%	52.9% (127)	91.0%	9.0%	55.8% (134)	17.2%	82.8%	84.6% (203)
Total	52.9% (127)	47.1% (113)		91.7% (220)	8.3% (20)		22.9% (55)	77.1% (185)	

A similar relationship was found with incorrect CDS which failed to alert the prescribing error. Omission errors were significantly higher when the drug reference for the medicine with the

prescribing error was not accessed in both low (Fisher's exact test, p=.002, n = 120) and high-complexity conditions (Fisher's exact test, p=.001, n = 120).

For correct CDS, there was no relationship between accessing the relevant drug reference and omission errors, as would be expected for correctly alerted prescribing errors (Table 1; Fisher's exact tests: Low-complexity, p=.731, n = 120; High-complexity, p=1, n = 120).

Over all conditions, 35% of participants in the control and 46% of participants in the incorrect CDS conditions made omission errors despite accessing the reference necessary to identify the error.

Clinical decision support reduced participants' access of drug references for medicines with prescribing errors.

Significantly fewer participants accessed drug references for medicines containing prescribing errors with incorrect compared to no CDS (control; McNemar tests: Low-complexity p<.001, n=120; High-complexity, p<.001, n=120). However, there was no difference in access between correct and no CDS (control; McNemar tests: Low complexity, p=.169, n=120; High complexity, p=.665, n=120).

Commission errors were higher when drug references relevant to false-positive alerts were not accessed.

False-positive alerts were more likely to lead to commission errors if the drug reference for the medicine triggering the alert was not accessed (Table 2; low-complexity, χ^2 (1, n = 116) = 16.673, p<.001, phi = -.379; high-complexity, χ^2 (1, n = 111) = 18.690, p < .001, phi = -.410.) Even when the relevant reference was consulted, 45.9% of participants across all conditions went on to make a commission error despite accessing references contradicting the alert.

Table 2 Percentage (number) of participants who accessed the drug reference relevant to the false-positive alert from incorrect CDS by whether a commission error was made.
Note: Includes only scenarios in which false-positive alerts were displayed.

	No error	Commission Error	Total
Low Complexity			
Accessed	48.4%	51.6%	53.4% (62)
Not accessed	13.0%	87.0%	46.6% (54)
Total	31.9% (37)	68.1% (79)	
High Complexity			
Accessed	61.2%	38.8%	44.1% (49)
Not accessed	21.0%	79.0%	55.9% (62)
Total	38.7% (43)	61.3% (68)	
Total			
Accessed	54.1%	45.9%	48.9% (111)
Not accessed	17.2%	82.8%	51.1% (116)
Total	35.2% (80)	64.8% (147)	

Task complexity did not affect access of drug references relevant to errors.

There was no difference in the proportion of participants who accessed drug references for medicines with prescribing errors (opportunities for omission errors) between the low- and high-complexity scenarios (McNemar tests: Control, p=.071, n=120; Correct CDS, p=.665, n=120; Incorrect CDS, p=.405, n=120.) Similarly, there was no difference in participants accessing drug references relevant to false-positive alerts (opportunities for commission errors) between the low- and high complexity scenarios (McNemar's test: Incorrect CDS, p=.117, n=108.)

Multilevel analysis of view time percentages

The multilevel analysis focused on the 466 scenarios (64.7%) in which drug references were accessed. View time percentage could not be calculated in one hundred scenarios (21.5% of these) where: task time was not recorded due to a software issue (n=93), outliers for task time (n=9) and view time (n=1) were removed, or view time data was missing (n=6). Several scenarios were affected by multiple issues. View time percentage was calculated for the remaining 366 scenarios (78.5%) and included in the model. With no systematic differences detected in the missing data; they were treated as being random.

Thirteen models were evaluated (appendix A) and from these four fixed effects were found to significantly contribute to the fit of a multilevel model and were included in the final model. The significance of fixed effects (predictors in the model) are reported in Table 3, and the model coefficients are presented in Appendix B. The comparison of effects is reported based on the estimated marginal means computed by the model. Significance probabilities have been adjusted for multiple comparisons using the Bonferroni correction.[37] The final model was significantly better than the intercepts only model, $\chi^2(7)=132.867$, p<.001. The intraclass correlation coefficient (ICC) was .23, indicating that 23% of the variance in verification was attributable to variation between participants, supporting the conduct of a multilevel analysis.[38, 39] The model residuals were normally distributed.

Table 3 Significance of fixed effects in the multilevel model of view time percentage.
*** Indicates significant effect (p<.05)**

	df	*F*	*p*
Intercept	1, 244.483	317.245	**<.001***
Task complexity Low complexity, high complexity	1, 302.436	105.383	**<.001***
Quality of decision support Correct CDS, incorrect CDS, control (No CDS)	2, 335.743	10.443	**<.001***
Omission error Omission error, no omission error	1, 361.914	4.498	**.035***
Quality of decision support * commission error	3, 346.223	2.712	**.045***

Participants who made omission errors spent significantly smaller percentage of task time viewing drug references (M = 24.7%, 95%CI [21.1%, 28.2%]) than those who did not make errors (M = 28.4%, 95%CI [25.1%, 31.6%]).

Similarly, participants who made commission errors with incorrect CDS spent significantly smaller percentage of task time viewing drug references (p=.018; M = 23.6%, 95%CI [20.0%, 27.2%]) than those who made no errors (M = 29.7%, 95%CI [25.6%, 33.6%]). This interaction occurs because only the incorrect CDS conditions displayed false-positive alerts that provided an opportunity for commission errors. There were no differences in the correct CDS (p=.977) or the control (p=.120) conditions.

View time percentage was significantly reduced by the provision of decision support (Figure 4). View time percentage was highest in the control condition which provided no decision support (M = 34%, 95%CI [29.7%, 39.9%]), and this was significantly higher than correct CDS (p<.001; M = 18.2%, 95%CI [12.7%, 23.8%]) and incorrect CDS (p=.012; M = 26.6%, 95%CI [23.7%, 29.5%]).

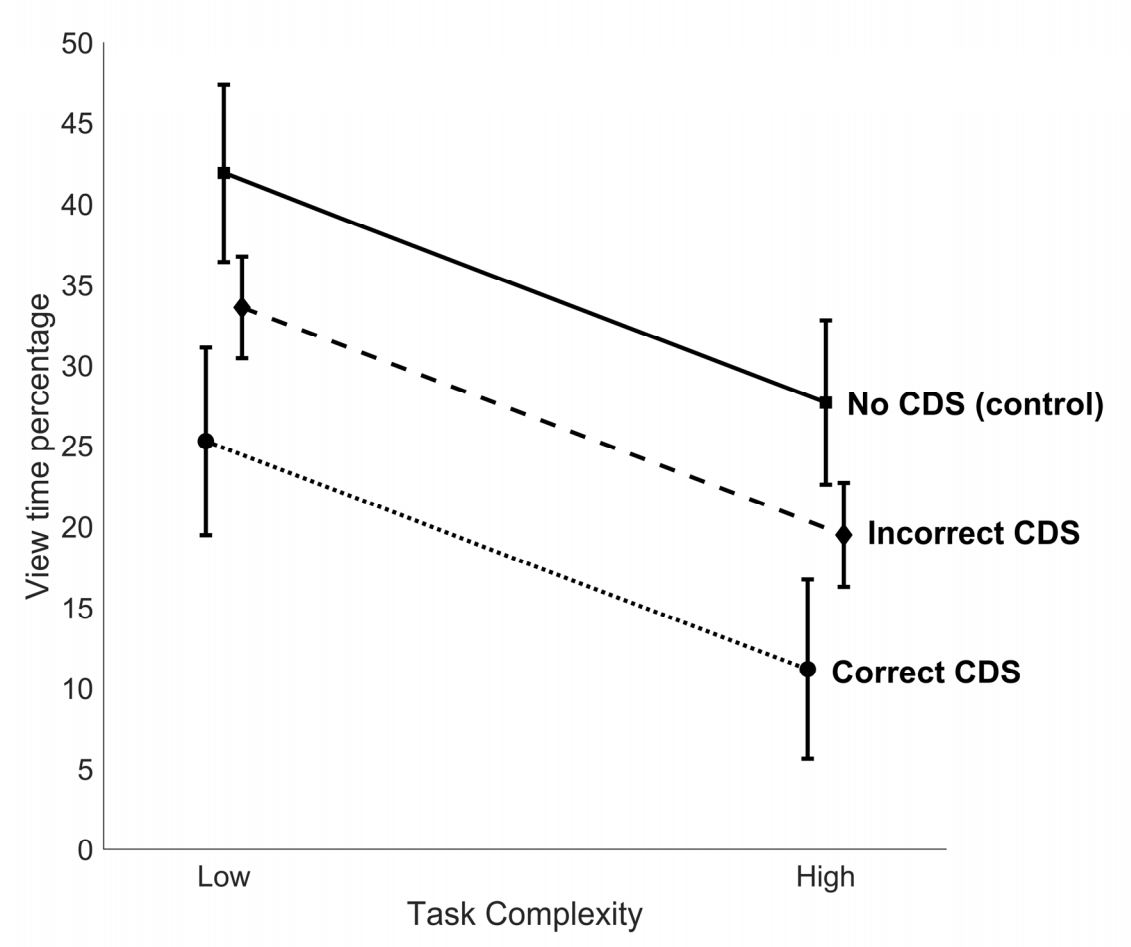

Figure 4 Estimated marginal means with 95% confidence interval (from the multilevel model) for view time percentage by task complexity and quality of decision support.

High task complexity significantly reduced view time percentage. Participants spent a significantly greater percentage of task time viewing drug references in low- (M = 33.6%, 95%CI [30.2%, 37.0%]) compared to high-complexity scenarios (M = 19.5%, 95%CI [16.4%, 22.5%]).

DISCUSSION

This experiment demonstrates firstly, that decreased verification, manifesting as either failure to access references or reduced view times as a percentage of task time, leads to increased omission and commission errors. Secondly, the presence of CDS decreases verification, and that decreased verification leads to increased omission and commission errors when CDS is incorrect.

We found that omission and commission errors increased when participants did not access relevant references (see Figure 5). Troublingly, some participants went on to make omission and commission errors despite accessing references containing information necessary to avoid those errors. Prior studies have reported a similar 'looking-but-not-seeing' or 'inattentional blindness,' [13, 15, 21] which describes how people may fail to perceive something in plain sight because they are not attending to it.[40] Consequently, accessing the relevant references did not guarantee errors were detected, but failure to do so made errors more likely.

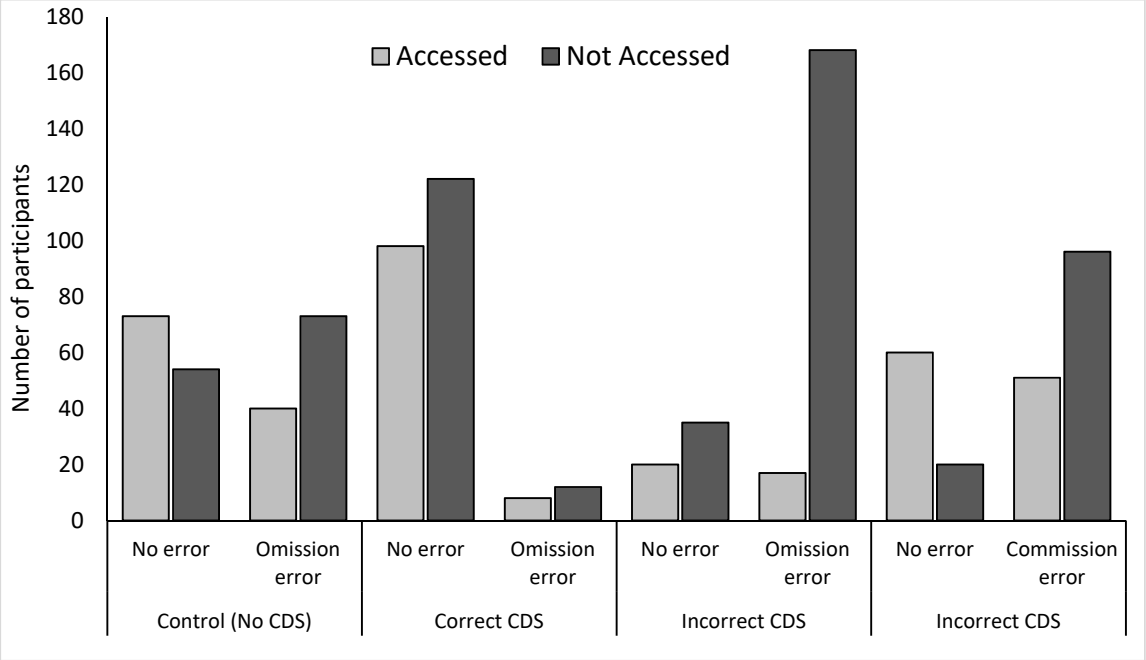

Figure 5 The number of participants who made errors by quality of CDS and whether the relevant drug reference was accessed.
Summarises the data presented in Tables 1 and 2, aggregating the low- and high-complexity conditions.

Seeking further insight into why accessing relevant references avoided some but not all errors, we analysed view time percentages. We found participants who avoided errors spent a significantly greater percentage of task time viewing references than those who made errors. Together the access and view time percentages results suggest: (1) verification should not be viewed as all-or-nothing,

but rather on a continuum of adequacy or vigilance, and (2) greater verification can reduce both omission and commission errors.

Clinical decision support reduced verification

We reported finding evidence of automation bias in this experiment; participants made significantly more omission and commission errors when provided with incorrect CDS compared to when they had no CDS.[17] The risk posed by automation bias is that CDS becomes a replacement for, rather than a supplement to, clinicians' efforts in error detection. The analysis of verification behaviour provides some support for the idea of CDS replacing participants' error detection efforts. A significantly smaller percentage of task time was spent viewing references in CDS-assisted compared to unassisted conditions (see Figure 4). This reduction in verification was associated with increased errors. It is very likely this relationship is causal with reduced verification impeding the discovery of errors.

Furthermore, when CDS was incorrect, participants who made omission or commission errors spent a smaller percentage of task time viewing references than those who did not make errors. This is consistent with prior automation bias research, which mostly employed aviation and process control tasks.[13-15, 19-22] The present study confirms this association extends to the detection of prescribing errors assisted by CDS medication alerts.

Manzey, et al. [13] suggest the looking-but-not-seeing effect, whereby participants made errors despite viewing information that could have prevented them, represents an automation bias induced withdrawal of cognitive resources for processing verification information. Therefore, while the necessary information was accessed, it was not processed in a way that enabled errors to be recognised. Our analysis of participants' cognitive load, reported separately, provides support for this. Participants who made omission errors allocated fewer cognitive resources to the task than those who did not.[28] Curiously, there was no difference for commission errors. The present findings suggest, in addition to reduced processing, there may also be reduced acquisition of information.

This is consistent with a *cognitive miser* view of automation bias,[16, 28] that people prefer adequate, faster and less effortful ways of thinking, rather than engaging in more accurate, but slower and more effortful thinking.[41] These findings also support Mosier and Skitka's description of automation bias as the use of automation as a heuristic,[16] with CDS appearing to be used as a shortcut in place of verification.

The same cognitive miser profile could also be found in participants who made errors in the control condition but to a significantly lesser extent. This may indicate the presence of other factors which trigger reduced verification in addition to automation bias.

Less verification in high complexity

High-complexity scenarios asked participants to prescribe five more medications, just over two and a half times the number requested in low-complexity scenarios. We expected the time to enter prescriptions into the e-prescribing system would increase as a function of the number of medications prescribed. Likewise, drug reference view time was expected to increase with the number of prescriptions and drug references that could be viewed. While there were no differences in access of relevant drug references as complexity increased from low to high, the view time percentage was significantly lower. The reduction in the percentage of task time viewing references could represent participants' efforts to manage the increased workload created by needing to verify more information in high-complexity scenarios. Despite this, we have previously reported that high task complexity did not increase automation bias errors.[17] This is puzzling, especially in light of present findings that high task complexity reduced verification, suggesting it may be a risk factor for automation bias. It is possible that participants' verification efforts were more sensitive to task complexity than errors, with both low- and high-complexity conditions exhibiting automation bias errors to a similar extent. If task complexity is a risk factor for automation bias, then both complexity conditions likely exceeded the threshold at which it presents. More research is needed to fully understand the relationship between task complexity and errors.

Implications

These findings highlight the importance of verification in preventing prescribing errors and may be generalisable to other forms of CDS. When prescribing is assisted by CDS medication alerts, verification provides the crucial means to differentiate between correct and incorrect CDS. However, the very presence of CDS is likely to exacerbate the problem, contributing to decreased verification, which, in turn, impedes the discovery of errors when CDS fails. This is the risk and challenge of automation bias. High task complexity further complicated matters, appearing to place downward pressure on verification, although the link between complexity and errors remains unclear. Improving the reliability and accuracy of CDS can reduce opportunities for error. However high-reliability automation is known to increase the rate of automation bias.[25] This risks clinicians being less able to detect CDS failures when they occur.

The challenge for designers and users of CDS is to ensure appropriate verification in circumstances that may promote decreased verification. To date, automation bias has proven stubbornly resistant to attempts to mitigate its effects,[23] including interventions which prompted users to verify.[42]

While our findings describe how CDS changed the access of references and view time percentages, little is known about what factors prompt clinicians to verify, the information sought and how they go about verifying, including the assessment of information and resolution of potential conflicts between different information sources. More research is needed in this area and how to best assist

clinicians with effective verification. Such efforts need to focus on how to best incorporate verification information into workflows, presenting only relevant information when, where and in the form it is needed. The challenge is to do this in a way that minimally impacts workload, doesn't overwhelm clinicians with too much information and maximises efficiency when CDS is correct.

Ultimately, clinicians need to be mindful that CDS can and does fail,[7-9] and that when it does, verification is the primary means to avoid errors. While it is impractical and undesirable to verify all prescriptions, clinicians would be well advised to verify whenever they suspect medication safety issues, even in the absence of medication alerts. It would also be prudent when prescribing unfamiliar or little-used medicines or for unfamiliar issues.

Limitations

This experiment was subject to several limitations. The use of medical students provided a necessary control for knowledge and experience of prescribing. This provides an indication of verification behaviour by junior medical officers entering practice but may have limited generalisability to more experienced clinicians. Clinician knowledge is likely to play an important role in verification but exceeds the scope of the present study. Likewise, the completeness of knowledge will also be an important consideration, for example, a clinician may know a medicine's contraindications for conditions, but not know all its possible adverse drug interactions.

Replication of our study with other cohorts, including more experienced clinicians, and clinicians operating in different clinical contexts would need to be undertaken. The evidence for the presence of similar verification results in other non-healthcare settings [13-15, 19-22] suggest however that these results are indeed generalisable to clinical decision making assisted by CDS.

Other factors that are likely to impact verification include the design and accuracy of CDS, and the accessibility of verification information. Further research identifying the relative contributions of such factors would be informative for developing mitigations.

Participants were not subjected to experimentally imposed time constraints or required to manage competing demands for their attention that clinicians would ordinarily experience in clinical practice.

Finally, the inclusion of conditions designed to elicit both omission and commission errors in the same condition means we cannot fully differentiate the effects of verification for each error type.

CONCLUSION

This is the first study to test the relationship between verification behaviours and the detection of prescribing errors, with and without CDS medication alerts. Increased verification was associated with increased detection of errors, while the presence of CDS and high task complexity reduced verification.

These findings demonstrate the importance of verification in avoiding prescribing and automation bias errors. CDS can alert clinicians to errors that may have been inadvertently missed, however they are not perfectly sensitive and specific. Clinicians should allow CDS to function as an additional layer of defence, but should not rely on it if they suspect a medication safety issue as it cannot replace the clinician's own expertise and clinical judgment.

CLINICAL RELEVANCE STATEMENT

Verification of CDS provides one means to avoid prescribing errors and is especially prudent when prescribing unfamiliar or little-used medicines or for unfamiliar issues. Clinical decision support medication alerts can help prevent prescribing errors, but CDS is imperfect and can be incorrect. The presence of CDS appears to reduce verification efforts and when CDS are incorrect reduced verification is associated with prescribing errors.

MULTIPLE CHOICE QUESTIONS

Question: What strategy can be used to reduce prescribing errors when using clinical decision support medication alerts?

a) Improve the accuracy of CDS medication alerts.
b) Verifying medication alerts, or their absence, with a gold standard, evidence-based drug reference.
c) Introduce messages into CDS systems that prompt clinicians to verify prescriptions.
d) Phase out CDS medication alerts.

Correct answer: The correct answer is b. Verifying medication alerts, or their absence, with a gold standard, evidence-based drug reference. Our results found that when CDS was incorrect greater verification was associated with reduced prescribing errors.

CDS medication alerts have been shown to reduce prescribing errors (not option d), but they introduce a risk of over-reliance. While, improving CDS accuracy would reduce opportunities for errors from over-reliance, perfectly sensitive and specific CDS are likely unattainable. Additionally, highly accurate decision support increases the rate of automation bias errors (not option a). Automation bias has proven stubbornly resistant to mitigations including prompting users to verify (not option c).

Question: When is verification of clinical decision support medication alerts, or their absence, especially prudent?

a) When prescribing unfamiliar or little-used medicines.
b) When prescribing for unfamiliar problems.
c) When a medication safety issue, such as contraindication, is suspected.
d) All of the above.

Correct answer: The correct answer is d. All of the above.

Clinicians will be familiar with and have a good knowledge of the medicines they frequently prescribe for commonly encountered issues. However, when prescribing unfamiliar or little-used medicines or prescribing for unfamiliar issues, clinicians may have gaps in knowledge and rely more heavily on CDS. If CDS is incorrect, however there is a risk of omission or commission errors occurring. In general, it is prudent for clinicians to verify computer-generated alerts, or their absence, if they suspect there is a risk of a prescribing error.

LIST OF ABBREVIATIONS

CDS: Clinical Decision Support

e-prescribing: Electronic Prescribing

FUNDING STATEMENT

This research was supported by a Doctoral Scholarship for David Lyell provided by the HCF Research Foundation.

COMPETING INTERESTS STATEMENT

The authors have no competing interests to declare.

CONTRIBUTORSHIP STATEMENT

David Lyell (DL) conceived this research and designed and conducted the study with guidance and under the supervision of Enrico Coiera (EC) and Farah Magrabi (FM).

DL drafted the manuscript with input from all authors. All authors provided revisions for intellectual content. All authors have approved the final manuscript.

ACKNOWLEDGMENTS

This research was supported by a doctoral scholarship for David Lyell provided by the HCF Research Foundation, Sydney, Australia. We acknowledge the contributions of Magdalena Z. Raban, L. G. Pont, Richard O. Day, Melissa T. Baysari, Vitaliy Kim, Jingbo Liu, Peter Petocz, Thierry Wendling, Robin Butterfield, Monish Maharaj, and Rhonda Siu, as well as the medical students who participated in this study.

ETHICS APPROVAL

The research was conducted in accordance with protocols approved by the Macquarie University Human Research Ethics Committee (5201401029) and the University of New South Wales Human Research Ethics Advisory Panel (2014-7-32).

REFERENCES

1. Thomsen LA, Winterstein AG, S⃝ndergaard B, Haugb⃝lle LS, Melander A. Systematic Review of the Incidence and Characteristics of Preventable Adverse Drug Events in Ambulatory Care. *Annals of Pharmacotherapy* 2007;**41**(9):1411-26 doi: 10.1345/aph.1H658
2. Tully MP, Ashcroft DM, Dornan T, Lewis PJ, Taylor D, Wass V. The Causes of and Factors Associated with Prescribing Errors in Hospital Inpatients. *Drug Safety* 2009;**32**(10):819-36 doi: 10.2165/11316560-000000000-00000
3. Wolfstadt JI, Gurwitz JH, Field TS, Lee M, Kalkar S, Wu W, Rochon PA. The Effect of Computerized Physician Order Entry with Clinical Decision Support on the Rates of Adverse Drug Events: A Systematic Review. *Journal of General Internal Medicine* 2008;**23**(4):451-58 doi: 10.1007/s11606-008-0504-5
4. Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The Effect of Electronic Prescribing on Medication Errors and Adverse Drug Events: A Systematic Review. *Journal of the American Medical Informatics Association* 2008;**15**(5):585-600 doi: 10.1197/jamia.M2667
5. van Rosse F, Maat B, Rademaker CMA, van Vught AJ, Egberts ACG, Bollen CW. The Effect of Computerized Physician Order Entry on Medication Prescription Errors and Clinical Outcome in Pediatric and Intensive Care: A Systematic Review. *Pediatrics* 2009;**123**(4):1184-90 doi: 10.1542/peds.2008-1494
6. Sweidan M, Williamson M, Reeve JF, Harvey K, O'Neill JA, Schattner P, Snowdon T. Evaluation of features to support safety and quality in general practice clinical software. *BMC Medical Informatics and Decision Making* 2011;**11**(1):1-8 doi: 10.1186/1472-6947-11-27
7. Wright A, Ai A, Ash J, Wiesen JF, Hickman T-TT, Aaron S, McEvoy D, Borkowsky S, Dissanayake PI, Embi P, Galanter W, Harper J, Kassakian SZ, Ramoni R, Schreiber R, Sirajuddin A, Bates DW, Sittig DF. Clinical decision support alert malfunctions: analysis and empirically derived taxonomy. *Journal of the American Medical Informatics Association* 2017;**25**(5):496–506 doi: 10.1093/jamia/ocx106
8. Wright A, Hickman T-TT, McEvoy D, Aaron S, Ai A, Andersen JM, Hussain S, Ramoni R, Fiskio J, Sittig DF, Bates DW. Analysis of clinical decision support system malfunctions: a case series and survey. *Journal of the American Medical Informatics Association* 2016;**23**(6):1068-76 doi: 10.1093/jamia/ocw005
9. Kassakian SZ, Yackel TR, Gorman PN, Dorr DA. Clinical decisions support malfunctions in a commercial electronic health record. *Appl Clin Inform* 2017;**08**(03):910-23 doi: 10.4338/ACI-2017-01-RA-0006
10. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of Drug Safety Alerts in Computerized Physician Order Entry. *Journal of the American Medical Informatics Association* 2006;**13**(2):138-47 doi: 10.1197/jamia.M1809
11. Nanji KC, Slight SP, Seger DL, Cho I, Fiskio JM, Redden LM, Volk LA, Bates DW. Overrides of medication-related clinical decision support alerts in outpatients. *Journal of the American Medical Informatics Association* 2014;**21**(3):487-91 doi: 10.1136/amiajnl-2013-001813
12. Oxford English Dictionary. *"verification, n."*: Oxford University Press, June 2018.
13. Manzey D, Reichenbach J, Onnasch L. Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience. *Journal of Cognitive Engineering and Decision Making* 2012;**6**(1):57-87 doi: 10.1177/1555343411433844
14. Bahner J, Huper A-D, Manzey D. Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies* 2008;**66**(9):688-99 doi: 10.1016/j.ijhcs.2008.06.001
15. Bahner J, Elepfandt MF, Manzey D. Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2008 Sep 22-26; New York, NY, United states. Human Factors And Ergonomics Society.
16. Mosier KL, Skitka LJ. Human decision makers and automated decision aids: Made for each other. In: Parasuraman R, Mouloua M, eds. Automation and human performance: Theory and applications. Hillsdale, NJ, England: Lawrence Erlbaum Associates, 1996:201-20.

17. Lyell D, Magrabi F, Raban MZ, Pont LG, Baysari MT, Day RO, Coiera E. Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making* 2017;**17**(1):28 doi: 10.1186/s12911-017-0425-5
18. Mosier KL, Skitka LJ, Heers S, Burdick M. Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology* 1998;**8**(1):47-63 doi: 10.1207/s15327108ijap0801_3
19. Bagheri N, Jamieson GA. The impact of context-related reliability on automation failure detection and scanning behaviour. IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583); 2004 Oct 10-13.
20. Bagheri N, Jamieson GA. Considering subjective trust and monitoring behavior in assessing automation-induced "complacency". In: Vincenzi DA, Mouloua M, Hancock PA, eds. Human Performance, Situation Awareness and Automation: Current Research and Trends, Vol 2. Mahwah: Lawrence Erlbaum Associates, 2004:54-59.
21. Reichenbach J, Onnasch L, Manzey D. Misuse of automation: The impact of system experience on complacency and automation bias in interaction with automated aids. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2010 Sep 27 - Oct 1; San Francisco, CA, United states. Human Factors And Ergonomics Society.
22. Reichenbach J, Onnasch L, Manzey D. Human performance consequences of automated decision aids in states of sleep loss. *Human Factors* 2011;**53**(6):717-28 doi: 10.1177/0018720811418222
23. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association* 2017;**24**(2):423-31 doi: 10.1093/jamia/ocw105
24. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 2012;**19**(1):121-27 doi: 10.1136/amiajnl-2011-000089
25. Bailey NR, Scerbo MW. Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science* 2007;**8**(4):321-48 doi: 10.1080/14639220500535301
26. Povyakalo AA, Alberdi E, Strigini L, Ayton P. How to Discriminate between Computer-Aided and Computer-Hindered Decisions: A Case Study in Mammography. *Medical Decision Making* 2013;**33**(1):98-107 doi: 10.1177/0272989x12465490
27. Sweller J. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review* 2010;**22**(2):123-38
28. Lyell D, Magrabi F, Coiera E. The Effect of Cognitive Load and Task Complexity on Automation Bias in Electronic Prescribing. *Human Factors* 2018;**60**(7):1008 - 21 doi: 10.1177/0018720818781224
29. De Vries TPGM, Henning RH, Hogerzeil HV, Fresle DA. *Guide to good prescribing*. Geneva: World Health Organization, 1994.
30. Australian Medicines Handbook Pty Ltd. Australian Medicines Handbook 2015 (online). Secondary Australian Medicines Handbook 2015 (online) 2015 January. Retrieved from http://amhonline.amh.net.au/.
31. Day RO, Snowden L. Where to find information about drugs. *Australian Prescriber* 2016;**39**(3):88-95 doi: 10.18773/austprescr.2016.023
32. Sweller J. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* 1994;**4**(4):295-312 doi: http://dx.doi.org/10.1016/0959-4752(94)90003-5
33. Sweller J, Chandler P. Why Some Material Is Difficult to Learn. *Cognition and Instruction* 1994;**12**(3):185-233 doi: 10.1207/s1532690xci1203_1
34. Hoffman L, Rovine MJ. Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods* 2007;**39**(1):101-17 doi: 10.3758/bf03192848
35. Tabachnick BG. *Using multivariate statistics / Barbara G. Tabachnick, Linda S. Fidel*. 6th edition. ed: Boston Pearson, 2013.
36. Peugh JL. A practical guide to multilevel modeling. *Journal of School Psychology* 2010;**48**(1):85-112 doi: http://dx.doi.org/10.1016/j.jsp.2009.09.002

37. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995;**310**(6973):170 doi: 10.1136/bmj.310.6973.170
38. Twisk J. Applied Multilevel Analysis: A Practical Guide. Cambridge, UK: Cambridge University Press, 2006.
39. Hayes AF. A primer on multilevel modeling. *Human Communication Research* 2006;**32**(4):385-410
40. Mack A, Rock I. *Inattentional blindness*. Cambridge, MA: MIT Press, 1998.
41. Fiske ST, Taylor SE. *Social cognition*. New York: Random House, 1984.
42. Mosier KL, Skitka LJ, Dunbar M, McDonnell L. Aircrews and automation bias: The advantages of teamwork? *International Journal of Aviation Psychology* 2001;**11**(1):1-14 doi: 10.1207/s15327108ijap1101_1

Article IV: Appendices

APPENDIX A: SELECTION OF PREDICTORS FOR MULTILEVEL MODEL

Model Evaluated	Fixed effect removed	-2 Log Likelihood	Number of Parameters	Likelihood ratio test	Include in final model
Fixed intercept, Random intercept (participant), Complexity, CDS, Omission error, Commission error, Complexity * CDS, Complexity * Omission error, Complexity * Commission error, CDS * Omission error, CDS * Commission error, Omission error * Commission error		2969.117	17		
Fixed intercept, Random intercept (participant), Complexity, CDS, Omission error, Commission error, Complexity * CDS, Complexity * Omission error, CDS * Omission error, CDS * Commission error, Omission error * Commission error	Complexity * Commission error	2969.118	16	$\chi^2(1)=.001$ p=.975	Exclude
Fixed intercept, Random intercept (participant), Complexity, CDS, Omission error, Commission error, Complexity * CDS, Complexity * Omission error, CDS * Commission error, Omission error * Commission error	CDS * Omission error	2969.282	14	$\chi^2(2)=.164$ p=.921	Exclude
Fixed intercept, Random intercept (participant), Complexity, CDS, Omission error, Commission error, Complexity * Omission error, CDS * Commission error, Omission error * Commission error	Complexity * CDS	2969.829	12	$\chi^2(2)=.547$ p=.761	Exclude
Fixed intercept, Random intercept (participant), Complexity, CDS, Omission error, Commission error, CDS * Commission error, Omission error * Commission error	Complexity * Omission error	2969.847	11	$\chi^2(1)=.018$ p=.893	Exclude
Fixed intercept, Random intercept (participant), Complexity, CDS, Omission error, Commission error, CDS * Commission error	Omission error * Commission error	2970.154	10	$\chi^2(1)=.307$ p=.580	Exclude
Fixed intercept, Random intercept (participant), Complexity, CDS, Omission error, Commission error	CDS * Commission error	2976.517	8	$\chi^2(2)=6.363$ p=.042	Include
Fixed intercept, Random intercept (participant), Complexity, CDS, Omission error	Commission error	2978.200	7	$\chi^2(1)=1.683$ p=.195	Exclude
Fixed intercept, Random intercept (participant), Complexity, CDS	Omission error	2983.210	6	$\chi^2(1)=5.010$ p=.025	Include
Fixed intercept, Random intercept (participant), Complexity	CDS	3029.940	4	$\chi^2(2)=46.730$ p<.001	Include
[Intercepts only model] Fixed intercept, Random intercept (participant)	Complexity	3103.021	3	$\chi^2(1)=73.081$ p<.001	Include
Fixed intercept	Random intercept (participant)	3114.720	2	$\chi^2(1)=11.699$ p=.001	Include
[Null model]	Fixed intercept	3567.464	1	$\chi^2(1)=452.744$ p<.001	Include

Final model

Intercept for fixed effects, Random intercept: Participant (Covariance structure: Variance components), Task complexity (low, high), Quality of clinical decision support (correct, incorrect, control), Omission error (no, yes), Quality of clinical decision support * Commission error (no, yes), Residual.

10 parameters, -2 Log likelihood = 2970.154

APPENDIX B: MULTILEVEL MODEL COEFFICIENTS

	Coefficient	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	29.81	5.048	356.884	5.905	.000	19.88	39.74
Low task complexity	14.17	1.380	302.436	10.266	.000	11.45	16.89
High task complexity	.	.	.	.	.	.	.
Correct CDS	-20.42	7.071	343.683	-2.888	.004	-34.33	-6.51
Incorrect CDS	-15.18	5.202	340.806	-2.918	.004	-25.41	-4.95
Control (No CDS)	.	.	.	.	.	.	.
No omission error	3.69	1.742	361.914	2.121	.035	.27	7.12
Omission error	.	.	.	.	.	.	.
Correct CDS * No commission error	-.16	5.346	339.548	-.029	.977	-10.67	10.36
Correct CDS * Commission error	.	.	.	.	.	.	.
Incorrect CDS * No commission error	6.00	2.520	353.936	2.381	.018	1.04	10.96
Incorrect CDS * Commission error	.	.	.	.	.	.	.
Control (No CDS) * No commission error	-7.87	5.047	345.557	-1.560	.120	-17.80	2.05
Control (No CDS) * Commission error	.	.	.	.	.	.	.

. = parameter is redundant.

Chapter 5 summary

5.5 Effect of task complexity and clinical decision support on verification

5.5.1 Omission errors

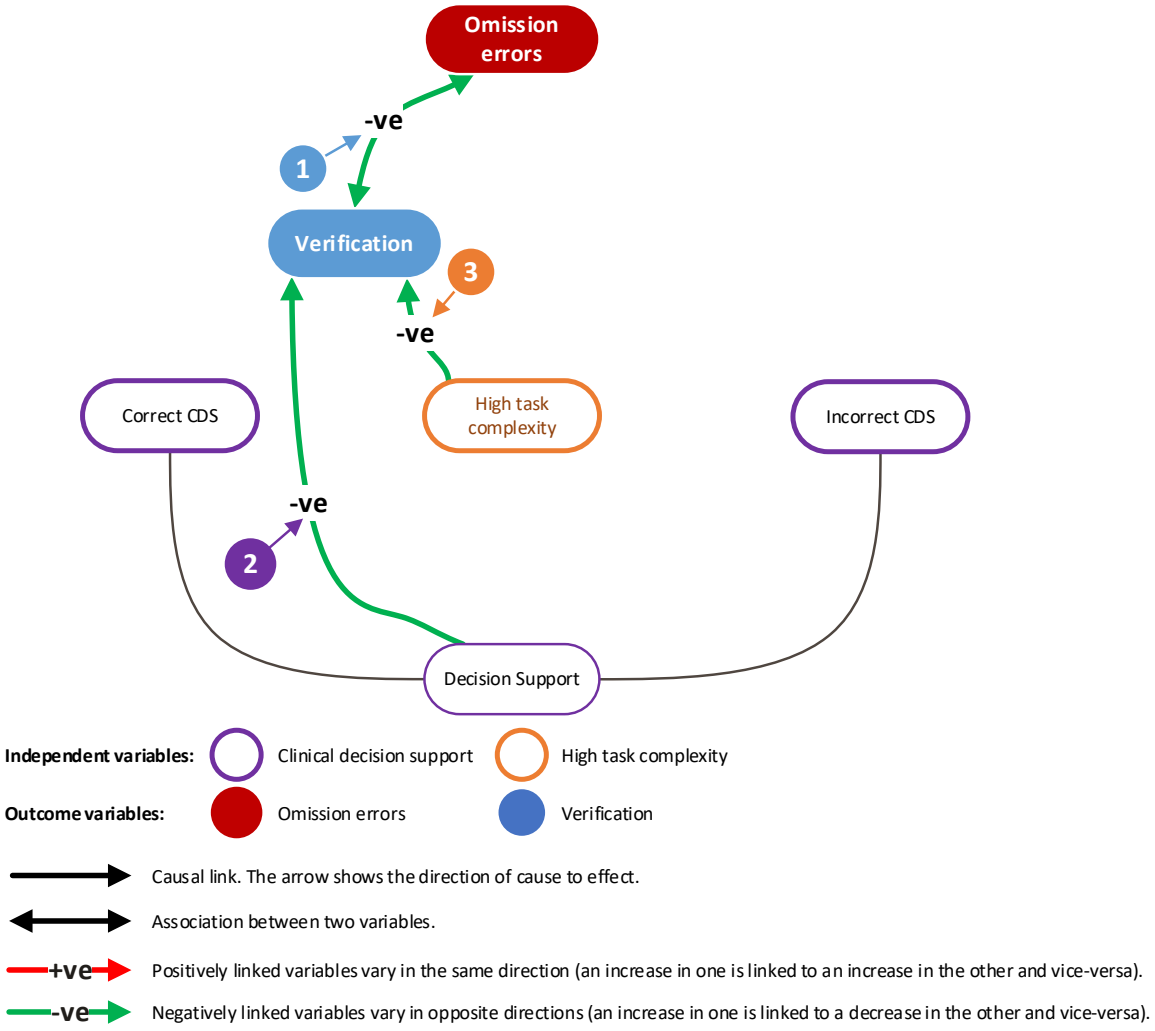

Figure 5-1 Effect of task complexity, clinical decision support and omission errors on verification

1 Increased verification was associated with decreased omission errors.

Participants who avoided omission errors spent 3.7% more task time engaged in verification than participants who made omission errors. See Table 3 in Article IV.

2 The presence of CDS reduced verification

Compared to when there was no CDS, correct CDS reduced verification time by 16.6% and incorrect CDS reduced verification time by 8.2%. See Figure 4 in Article IV.

3 Increasing task complexity decreased verification.

Increasing task complexity from low to high decreased verification time by 14.2%.

5.5.2 Commission errors

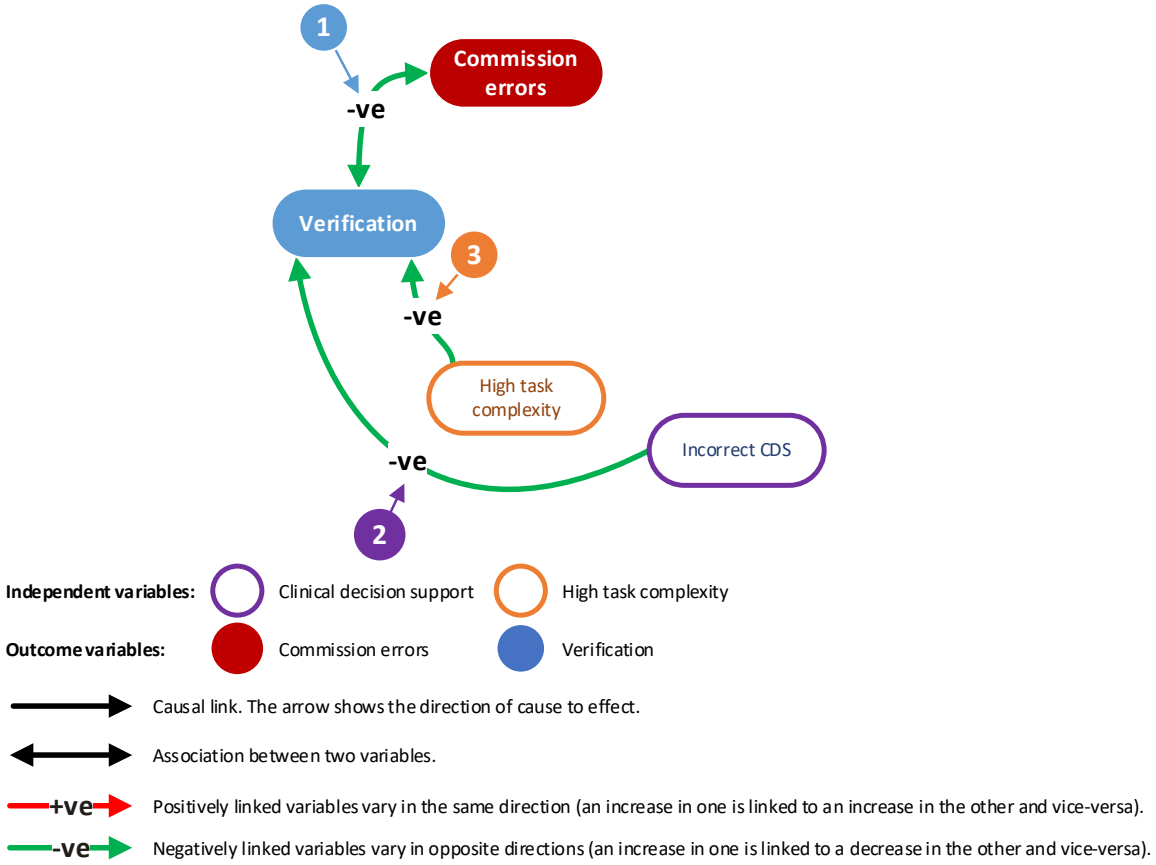

Figure 5-2 Effect of task complexity, clinical decision support and commission errors on verification

1 Increased verification was associated with decreased commission errors.

Participants who avoided commission errors from incorrect CDS spent 6% more task time engaged in verification than participants who made commission errors. See Table 3 in Article IV.

2 The presence of CDS reduced verification

Compared to when there was no CDS, correct CDS reduced verification time by 16.6% and incorrect CDS reduced verification time by 8.2%. See Figure 4 in Article IV.

3 Increasing task complexity decreased verification.

Increasing task complexity from low to high decreased verification time by 14.2%.

5.6 Chapter 5 References

1. Australian Medicines Handbook Pty Ltd. Australian Medicines Handbook 2015 (online). Secondary Australian Medicines Handbook 2015 (online) 2015 January. Retrieved from http://amhonline.amh.net.au/.

2. Lyell D, Magrabi F, Coiera E. Reduced Verification of Medication Alerts Increases Prescribing Errors. *Applied Clinical Informatics* 2019;**10**(01):066-76 doi: 10.1055/s-0038-1677009

6 Discussion

This thesis sought to study automation bias in the delivery of healthcare, focusing on the tasks, automation, and risk factors which are characteristic of healthcare settings and applications. In so doing, this thesis has contributed the first evidence of automation bias in the electronic prescription of medicines, a common task performed by clinicians that is assisted by clinical decision support. It has advanced greater insights into how people who suffer an automation bias are more likely to make errors, spend less time engaged in verification, and allocate fewer cognitive resources to their task. This thesis has contributed the first analysis of differences in cognitive load in terms of whether errors were made. It has also found that, while higher task complexity did not induce or increase automation bias errors, it did appear to adversely impact verification and cognitive load.

6.1 Contributions of this thesis

The systematic review [Article I; 1] identified the aspects of automation bias that are unique to healthcare, an area where further research is needed (aim 1). These included decision-making tasks which were classified as either diagnosis or treatment, where the decision maker was required to identify the current state of a system (or patient), identify the cause of a problem or decide how to best remedy the problem. All healthcare studies took place in a single task environment. This contrasts with non-healthcare studies predominantly from the heavily automated domains of aviation and process control that take place in multitask environments. The review found that, contrary to the prevailing view within the human factors literature, automation bias can present in a single task environment. Moreover, this appears to be associated with higher verification complexity, that is, the task complexity of verifying automation. Tasks with lower complexity only seemed to produce automation bias when combined with concurrent secondary tasks in a multi-task environment.

The systematic review initially identified two distinct bodies of work that were labelled based on the area of publication (healthcare and human factors). However, the analysis within the review centred on tasks. Accordingly, over the course of the research, the labelling of the differences reported in the review shifted to one based on tasks rather than areas of publication. This shift better captures the relationship between the two bodies of literature: human factors and ergonomics is a field of study and practice, while health information technology is an application domain. Indeed, the publication of article III [2] in the journal *Human Factors* exemplifies this point.

Similarly, the claim in the systematic review about the prevailing view among researchers publishing in the human factors and ergonomics literature may seem controversial. Nevertheless, it is based on

the prevalence of multitask studies reported within this literature as observed in the systematic review,[1] and a 2010 review by Parasuraman and Manzey.[3] The latter proposed an integrated model of automation bias and complacency premised on the notion that the phenomenon occurs "under conditions of multiple-task load, when manual tasks compete with the automated task for the operator's [limited] attention."[3] The prevalence of multitask studies is likely influenced by prior research. The first reported study of automation bias found no evidence in an air traffic control task performed in a single-task environment.[4] Parasuraman and his colleagues' hypothesis [5] was that the absence of an effect in that study may have arisen because: (1) participants were responsible for a single task, and (2) automation bias may be more likely when users perform multiple tasks. This hypothesis was confirmed in experiments finding evidence of automation bias in a system monitoring task in a multi-, but not in a single-task environment.[5, 6]

The experiment was the first to test for, and find evidence of, automation bias errors in e-prescribing, where participants were assisted by CDS alerts which warned them of potential errors in their prescriptions [aim 2; Article II; 7]. This represents a common use of a decision support system that is encountered in everyday clinical practice. The results demonstrated that, when prescribing with CDS, there is a risk of automation bias for junior medical officers who are commencing clinical practice. These findings of automation bias are likely to be generalisable to other forms of decision support among this cohort of users.

Additionally, the experiment set out to experimentally test three hypothesised causes of automation bias that constitute risk factors characteristic of healthcare settings. Specifically, it tested whether automation bias errors may be induced by: (1) high task complexity [aim 3; Article II; 7], (2) interruptions [aim 4; Article II; 7], and (3) high cognitive load [aim 5; Article III; 8]. Ultimately, no evidence was found to support any of these hypothesised causes.

While the task complexity manipulation failed to alter the rate of automation bias, the results provided an excellent view of automation bias occurring. This enabled an understanding of how automation bias impacted each of the outcome variables (aim 6). Automation bias increased omission and commission errors.[Article II; 7] Multilevel modelling revealed that reduced cognitive load was associated with omission errors,[Article III; 8] while reduced verification was associated with omission and commission errors.[Article IV; 2] These relationships are illustrated in Figure 6-1, which is similar to the hypothesised relationship presented in Figure 1-1, based on Mosier and Skitka's definition of automation bias.[9]

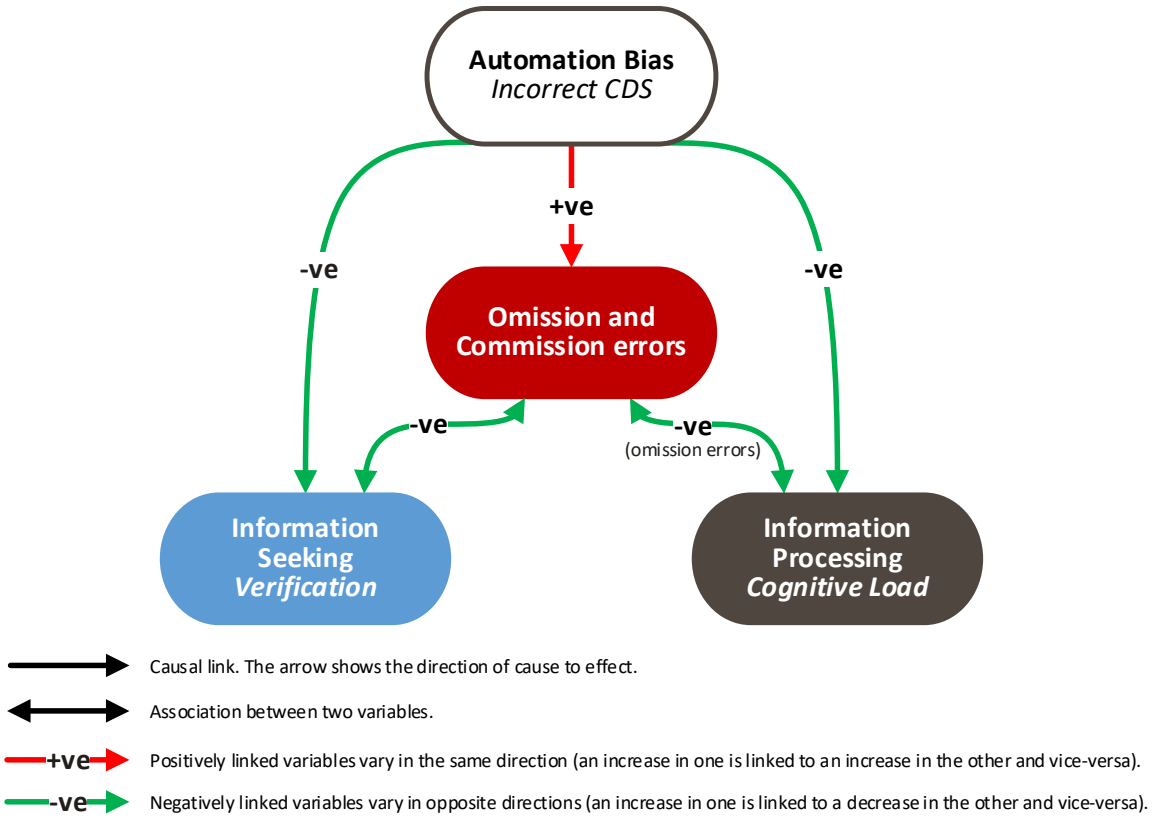

Figure 6-1 Relationship between automation bias, information processing, information seeking and errors

The design of the randomised controlled experiment demonstrated a robust and replicable method for the study of automation bias in healthcare settings which use decision support.[10] The presence of automation bias was established by comparing incorrect CDS which provided opportunities for omission and commission errors with a control condition where participants had to perform the task manually without decision support. The control condition provided baseline measurements for errors, cognitive load and verification to which the adverse impact of incorrect CDS could be compared.

Cognitive Load Theory was introduced as a framework for manipulating and testing the effect of task complexity and has proven invaluable in understanding how participants allocated cognitive resources to the e-prescribing task. Additionally, Cognitive Load Theory provides a large body of research on cognitive load effects which have been shown to impact learning and problem-solving outcomes in education.[11] Many effects focus on the reduction of extraneous cognitive load in educational applications; however, they could be equally applied to the design of e-prescribing systems and clinical decision support, and may assist in the development of enhanced verification for CDS.

Participants' access of drug references provided an observable measure of verification. It also provided insights into how participants' behaviour was changed by the presence and correctness of CDS, task complexity and whether errors were made.

The knowledge and experience of participants were controlled by recruiting students in the final two years of their medical degree. It was expected that, while participants would have all received training in the safe and rational use of medicines, their knowledge of, and expertise in, specific pharmacological therapies would be reasonably low. This provided two very important baselines for this study: (1) the measurement of cognitive load, and (2) the measurement of verification, both of which are dependent on existing knowledge or schema. This enabled between-subjects comparisons, such as exploring differences between participants who made errors and those who did not. However, some variation in expertise for specific medicines was to be expected.

Randomisation was essential to control for order or learning effects. Awareness of CDS failures may have altered participants' trust in CDS and therefore their reliance on, and compliance with, it. The randomisation of patient scenarios to experimental conditions was necessary to prevent differences in participants' knowledge of different medicines from influencing the results.

6.2 Automation bias effects model

A model of the effects of task complexity and quality of decision support on each dependent variable was progressively constructed for errors (chapter 3), intrinsic cognitive load (chapter 4) and verification (chapter 5). The model provides a graphical view of the relationships between the experimental variables.

6.2.1 Omission error model

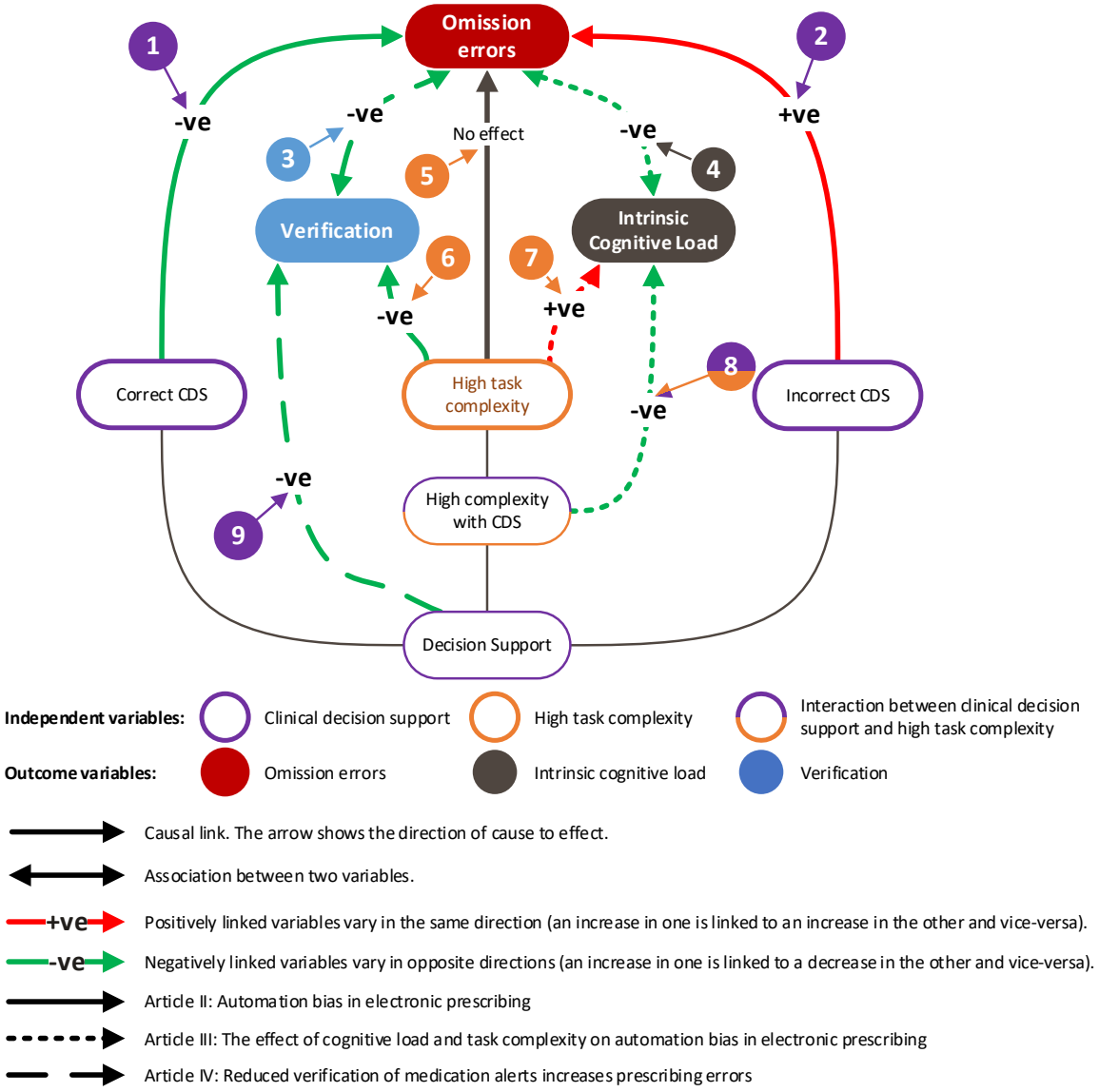

Figure 6-2 Automation bias omission error model

The omission error model (Figure 6-2) demonstrates that CDS provides both benefits and risks. Correct CDS reduces omission errors 1, while incorrect CDS increases them 2.

The increase in omission errors caused by incorrect CDS 2 represents automation bias omission errors. Omission errors were associated with reduced verification 3 and intrinsic cognitive load 4.

Increasing task complexity from low to high did not affect omission errors 5. High task complexity decreased verification 6 and increased intrinsic cognitive load 7.

In high complexity scenarios, the presence of CDS decreased intrinsic cognitive load 8. The presence of CDS reduced verification 9.

6.2.2 Commission error model

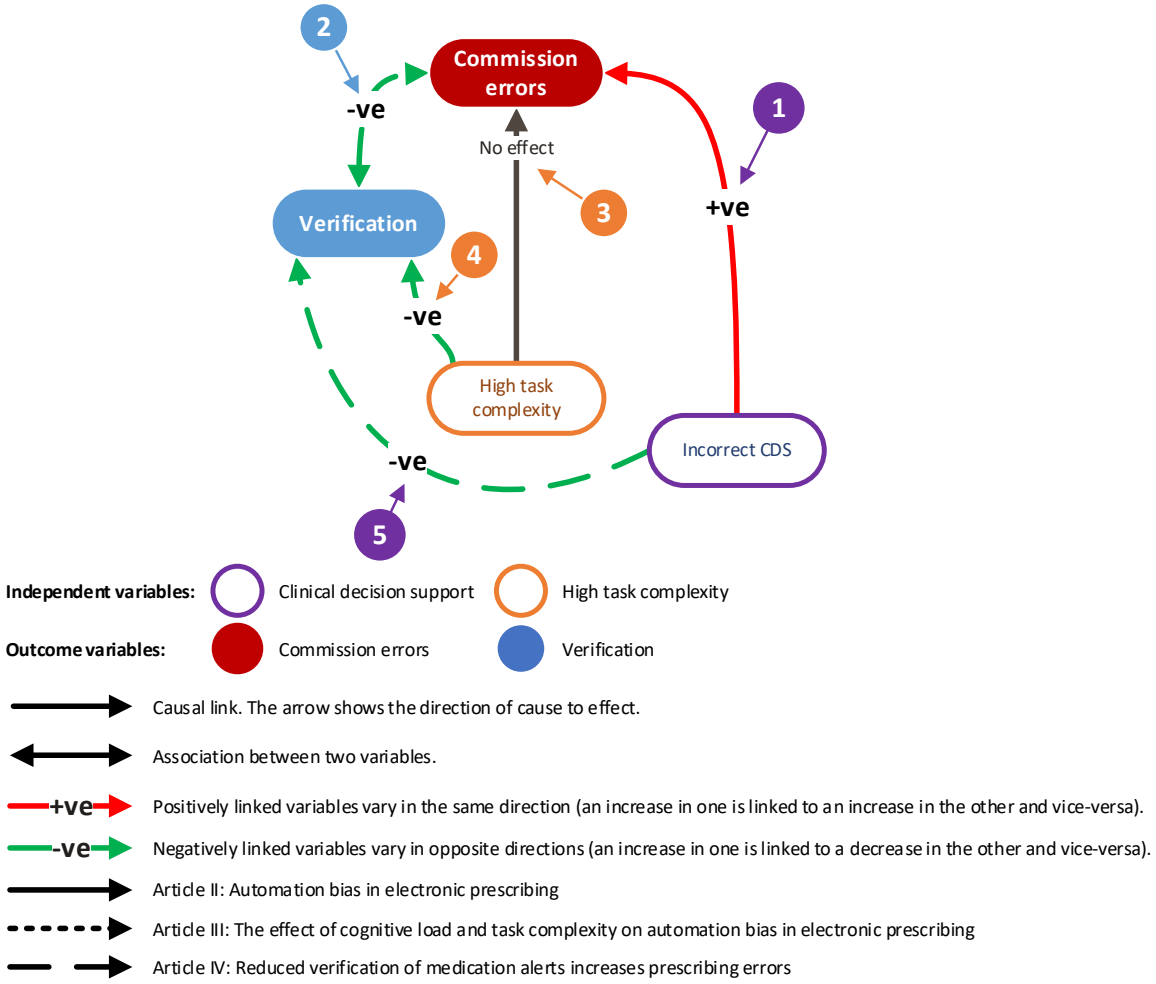

Figure 6-3 Automation bias commission error model

The commission error model (Figure 6-3) illustrates the effects of false-positive alerts from incorrect CDS.

Incorrect CDS increased commission errors (1). Commission errors were associated with reduced verification (2).

Increasing task complexity from low to high had no effect on commission errors (3), but decreased verification (4). The presence of CDS reduced verification (5).

The commission error model (Figure 6-3) differs from the omission error model (Figure 6-2) in two main ways: (1) commission errors were unaffected by Correct CDS, and (2) there was no association between commission errors and intrinsic cognitive load.

6.3 Decision support as a heuristic

Mosier and Skitka define automation bias as the tendency of people to use automated cues, such as advice and alerts from CDS as a "heuristic replacement for vigilant information seeking and processing."[9]

Humans tend to be *cognitive misers*, preferring to travel the path of least cognitive effort, seeking adequate decisions quickly and with minimal effort, rather than engaging in more accurate but slower and more effortful decision-making.[12] This is achieved by employing heuristics, which are simple rules or cognitive shortcuts people may use to make efficient decisions, especially in the face of uncertainty. Heuristics are frugal; they use some, but not all, of the available information.[13] Ideally, heuristics should provide quick and simple ways of making decisions that are sufficiently accurate most of the time [9] and are good enough for their purpose.

In this way, it is possible for clinicians to use clinical decision support alerts as a heuristic to quickly and effortlessly determine whether a prescription is safe. Indeed, in the e-prescribing experiment, the presence of CDS reduced both the time engaged in verification [Article IV; 2] and cognitive load when task complexity was high.[Article III; 8] Furthermore, it improved the detection of prescribing errors, but only when CDS was correct.[Article II; 7] This reduction in verification and cognitive load supports the notion that CDS was used as a heuristic in determining whether prescriptions were safe.

When CDS was correct, it proved to be an extremely effective heuristic. But, when incorrect, it led to error,[Article II; 7] demonstrating how heuristics become biases when they systematically and predictably lead to erroneous decisions. Participants who made omission and commission errors exhibited a cognitive miser profile, investing less time in verification,[Article IV; 2] and those who made omission errors invested fewer cognitive resources.[Article III; 8] While a similar profile was found for participants who made errors in control scenarios without CDS, this was to a significantly lesser extent. These findings provide support for the notion that CDS was relied on as a heuristic for detecting prescribing errors, with participants forgoing the information seeking and processing necessary to recognise prescribing errors and CDS failures.

More broadly, the relationship between automation bias errors and reduced verification has been well established in the research literature.[14-20] The cognitive load analysis presented in Article III [8] provided the first evidence of the relationship between omission errors and reduced allocation of cognitive resources.

The use of CDS as a heuristic may also explain the 'looking but not seeing' effect observed, whereby participants made errors despite accessing the information necessary to identify the error. This is similar to *inattentional blindness,* where people fail to see something in plain sight because their attention was not directed towards it.[21] Manzey, et al.,[18] suggest that automation bias may not

necessarily result in a complete neglect of verification, but rather in a withdrawal of attentional resources for processing looked-at verification information. This is supported by research that has found that participants who made commission errors were unable to correctly recall information they had accessed [18] or falsely recalled information which supported their agreement with incorrect decision support.[22] Some participants made commission errors despite having an awareness of contradictory verification information. This may indicate an active discounting of that information. However, this was rarer than incorrect recall.[18]

While the conclusions drawn by Manzey, et al.,[18] relate to commission errors, the experiment presented in this research suggests that their conclusions are also likely to be generalisable to omission errors. Although the experiment could not confirm whether participants were aware of contradictory information, reduced verification time and cognitive load support the idea of the withdrawal of cognitive resources. This is troubling; these participants appeared to go through the motions of verification, but without the sufficient cognitive engagement required for verification to be effective. For this reason, it is important to assess not only the access of verification information, but also the quality of verification, such as by testing the recall of key verification information [18, 22], cognitive load [8] and view time.[2]

To be useful, heuristics need to enable efficient decision-making that is good enough for their purpose. While the purpose of CDS used in e-prescribing is to prevent prescribing errors, what is *good enough* is determined by the risk of harm to patients. The presence of CDS certainly demonstrated the potential to enhance efficiency by reducing cognitive load and verification, but this was dependent on the quality of the decision support provided. Correct CDS resulted in substantial improvement. However, incorrect CDS meant that participants ordered prescriptions for the hypothetical patients which contained significant to potentially lethal errors and also that appropriate treatments were not prescribed. Given the risk of harm, there is also a risk that decisions will not be good enough when CDS is incorrect.

6.4 Task complexity

One primary aim of this research was to establish whether high task complexity induced or exacerbated automation bias (aim 3) and whether the effect of high task complexity might be explained by high cognitive load (aim 4). Ultimately, the analyses reported in Article II [7] and Article III [8] found no support for high task complexity or high cognitive load as a causal or aggravating factor that affected automation bias errors.

Interestingly, however, high task complexity did impact cognitive load and verification in ways that appeared to increase the likelihood of automation bias errors. Specifically, Article III [8] reported how the presence of CDS reduced cognitive load when task complexity was high. Article IV [2] described how both the presence of CDS and high task complexity reduced verification. While these findings

could indicate greater reliance on CDS when task complexity is high, high task complexity did not translate into higher error rates when assisted by incorrect CDS. This seems to indicate that cognitive load and verification were more sensitive to task complexity than automation bias errors.

Participants appeared to find both levels of task complexity to be difficult, with both low- and high-complexity conditions producing automation bias errors.[7] Consequently, if task complexity is a cause of automation bias, then it is likely that the low complexity task exceeded the threshold at which automation bias presents.[8] Indeed, there is some research that suggests that this threshold could be quite low. Harris and Goernert [23] observed that participants appeared to make minimal use of a decision aid and instead made decisions based on inspection of verification data when performing a simple task. This was attributed to the minimal reduction in workload offered by automation for a simple task. This suggests that the perceived benefits offered by automation, such as reduced workload,[23] cognitive load and verification time, and better performance, may be important contributors to automation bias. It also suggests that the benefits of automation may be greater under conditions of high task complexity. In multi-task environments, automation provides the benefit of allowing the reallocation of attention to other tasks.[3] This may explain the observation of the systematic review [Article I; 1] that automation bias presented in low-complexity multitasks, but not in single tasks.

The relationship between task complexity and automation bias appears to be more complicated than originally thought. There is a clear and consistent effect of high task complexity and the presence of CDS on cognitive load and verification. Therefore, task complexity does seem to have a role in automation bias; however, the full extent of this remains unknown. There is reasonable evidence to suggest that tasks may require a minimum level of complexity for automation bias to occur. It could be that once this minimum threshold of complexity is met, decision support provides sufficient benefit to the user to induce reliance. More research is needed to test these specific hypotheses.

6.5 Implications

Clinical decision support is a double-edged sword. Correct CDS can prevent harmful prescribing errors [24-26] and in the experiment proved extremely effective, reducing prescribing errors with less verification and cognitive load.[2, 7, 8] However, incorrect CDS caused errors that would not have otherwise occurred.

Automation bias only becomes problematic when CDS is incorrect, that is, in situations where overreliance can lead to omission or commission errors. CDS can be incorrect for diverse reasons: the fact that not all clinical problems are evaluated by CDS,[27] the occurrence of malfunctions,[28-30] and the frequent over-riding of alerts.[31, 32]

The naïve solution would be to eliminate incorrect CDS; however, perfectly accurate CDS is unlikely to be attainable, especially for complex clinical decision-making involving uncertainty and requiring expert judgement. More importantly, highly accurate decision support is known to increase the rate of automation bias,[5, 6, 14, 33-39] meaning that when incorrect CDS presents, it is even less likely to be detected. Consequently, rather than just focusing on CDS, the bias itself must also be tackled.

If omission and commission errors result from the difficulty of discriminating incorrect from correct CDS,[7] then verification is a promising albeit challenging mitigation to pursue. While, there is good evidence associating greater verification with reduced automation bias errors,[2, 14-20] the very presence of CDS causes a reduction in verification.[2] Furthermore, requiring complete verification of all CDS would negate many of the benefits offered by correct CDS.

Therefore, the challenge is to facilitate appropriate verification that will uncover incorrect CDS, without losing the benefits of correct CDS. Unfortunately, automation bias has proven stubbornly resistant to interventions aimed at mitigating its effects.[1] Such interventions include training on avoiding automation bias errors,[40] exposure to incorrect automation during training,[16] externally imposed accountability for performance,[22, 41] feedback on performance,[42] and even prompts to verify.[40]

The interface between the user and CDS provides an opportunity to facilitate appropriate verification. To be effective, mitigations need to incorporate verification information into clinical workflows and present relevant information at the time of decision-making. Interventions need to enable users to recognise incorrect CDS, but with minimal impact on workload and the avoidance of information overload.

More research will be needed to progress such efforts insofar as little is known about what prompts clinicians to verify CDS, how they go about verifying and what information they seek. Research exploring verification of CDS in clinical environments could provide valuable insights which could inform the design of verification systems. It will also be especially important to understand how people recognise incorrect CDS and how they resolve discrepancies between CDS and other information sources. Such insights would require laboratory experiments that expose clinicians to incorrect CDS, which would not be feasible in clinical settings.

A significant barrier to effective verification is the *looking but not seeing effect*;[2, 17-19] to date, this effect has been reported as an emergent finding of automation bias studies but has yet to be studied directly. Research that employs eye tracking and measures eye gaze dwell time could provide valuable insights by evaluating whether participants see critical verification information. Measurement of cognitive load could enable greater understanding of the relative contributions of adequate information searching and the withdrawal of cognitive resources.[18]

While the effect of task complexity on automation bias errors remains unresolved,[7, 33, 43, 44] high task complexity was shown to adversely impact cognitive load [8] and verification.[2] It is possible that verification complexity presented an obstacle for recognising incorrect CDS,[1] even though it did not translate into higher rates of automation bias.[7] Cognitive Load Theory provides a useful framework for assessing the cognitive demands of verification, as well as a body of well-studied *cognitive load effects* which can be leveraged to reduce cognitive load.[11] Future research could establish the role of verification complexity and whether reducing complexity and cognitive load may prevent both automation bias errors and *looking but not seeing* effects.

Automation bias becomes a risk when CDS becomes a replacement for the clinician's efforts to ensure the safe prescribing of medicines. This results in the subversion of CDS's proper function, namely, to provide an extra layer of defence against prescribing errors. Ultimately, clinicians are responsible for the treatments they prescribe, and therefore should be mindful that CDS is imperfect and can be incorrect.[27-32] Verification is the primary means of avoiding prescribing errors and distinguishing incorrect from correct CDS. It would be impractical to expect clinicians to verify all of their prescriptions. Nevertheless, they would be well-advised to investigate and verify any prescription they suspect may pose a risk of patient harm, even in the absence of CDS medication alerts. Verification would also be especially prudent when prescribing unfamiliar or little-used medications and for unfamiliar issues.

6.6 Limitations

Studying automation bias within an experimental context also presents some challenges. Chief among these is the relationship between the participant and automation or decision support. Highly reliable automation leads to higher rates of automation bias.[5, 6, 14, 33-39]. This is likely due to greater reliability engendering greater trust, where greater trust has been linked to automation bias omission errors.[15, 33]. By contrast, experiencing automation failures lowers trust.[19] As such, experience over time with automation may help to calibrate users' trust in it, which, in turn, may affect their reliance on, and compliance with, automation.[3]

For this experiment, there was limited capacity to manipulate the calibration of trust between the participants and CDS. They were introduced to CDS of which they had no prior knowledge or experience, and then exposed to incorrect CDS. The CDS manipulations may affect the calibration of trust and, in turn, reliance on CDS. For this reason, it would be unfeasible to establish the real world incidence or prevalence of automation bias errors experimentally. It is likely that any observed rate of automation bias will be specific to the decision aids and context used in a particular study.

Recruiting medical students for this research involved a compromise. On the one hand, recruiting medical students allowed for more control; it ensured that all participants had an equivalent amount of knowledge and experience which was important for comparing cognitive load and verification. On

the other hand, the results of the research have limited generalisability. They may be indicative of junior medical officers entering clinical practice, but not of experienced clinicians with a greater knowledge of pharmacological therapies and greater experience with e-prescribing systems.

Additionally, CDS in current e-prescribing systems tend to display a large number of clinically insignificant alerts. These alerts, while technically correct, relate to issues which may not be clinically significant insofar as they are interruptive but do not require action from the clinician. Consequently, these alerts are false-positives and CDS is therefore incorrect for that decision. This is evidenced by the high override rate of alerts.[32] Troublingly, clinicians can become desensitised to alerts, thereby introducing the risk that alerts for clinically significant issues might also be overridden. This phenomenon is known as alert fatigue.[32, 45]

Differentiating clinically significant from insignificant alerts is aided by verification, but also requires clinical judgement to evaluate whether the risks identified are acceptable and outweighed by the benefits of a particular treatment. Therefore, it is likely that recognising clinically insignificant false-positives may differ from recognising factually incorrect alerts. The experiment examined factually incorrect false-positives to control for errors arising from the clinical judgement of medication risks and benefits in a cohort of medical students with little prescribing experience. Nevertheless, addressing alert fatigue that stems from low specificity CDS is also crucial for effective CDS. Further research is needed to determine the biasing effect of clinically insignificant alerts from high specificity CDS in relation to commission errors.

The inclusion of opportunities for both omission and commission errors in the same conditions limited the ability to fully differentiate the cognitive load and verification effects for each error type. This may be important in light of evidence that people are differently affected by false-negative and false-positive alerts.[17] It is very likely that an alert, regardless of its correctness, provides a cue within the decision-making environment. For reasonably specific CDS, such a cue may act as a prompt for users to further investigate or verify the alert, thereby initiating actions that could increase the probability that false-positives are detected. If so, then commission errors would be less likely than omission errors where there is no such cue. There were more omission than commission errors in the experiment, although this was not statistically significant. However, this should be studied further and could have implications for understanding trade-offs between the sensitivity and specificity of CDS.

Additionally, the theoretical treatment of automation bias in this thesis has followed the work of prior researchers who have written on the topic, most notably Mosier and Skitka's view of automation bias as a heuristic [9] and the characterisation of errors despite verification as being like inattentional blindness.[18] This research introduced Cognitive Load Theory [11] as a framework to explore whether task complexity might explain why some single-tasks resulted in automation bias while others did not. It also enabled new insights into the association between automation bias and the reduced allocation

of cognitive resources and how this might contribute to automation bias and the failure to recognise errors despite verifying. There are other theoretical frameworks which have not previously been explored in the automation bias literature that may offer new insights or different perspectives on the relationship between automation bias and how people allocate cognitive and other resources to tasks. Future research could review the available evidence, including the findings presented in this research, and evaluate coherence with existing and new theoretical frameworks, including Dual-Process Theory[46] and the Soft Constraints Hypothesis.[47, 48]

6.7 Conclusion

This thesis set out to study automation bias in healthcare, focusing in particular on the tasks, automation and risk factors which may be unique to, or feature prominently in, healthcare applications. Significant evidence of automation bias was found in an e-prescribing task that was assisted by CDS medication alerts, which is a common clinical decision-making task supported by a frequently-encountered form of decision support. Participants made omission errors by failing to detect prescribing errors that CDS did not alert. They also made commission errors when they accepted factually incorrect false-positive alerts.

While not supporting high task complexity as a cause of automation bias, the experiment did provide an excellent view of automation bias occurring. Omission errors were associated with less verification time and lower cognitive load, while commission errors were associated with less verification. The results also revealed the tendency of participants, when assisted by CDS, to reduce verification and reduce the cognitive resources allocated to high complexity tasks.

These findings support a cognitive miser view of automation bias, suggesting that CDS medication alerts were used as a heuristic shortcut in lieu of deliberate information seeking and processing. This produces a state that is highly likely to compromise the effectiveness of verification. Although causality cannot be inferred from these findings, it is highly likely that when people suffer an automation bias, they reduce both verification behaviours and the cognitive resources allocated to processing looked-at information. Ultimately, this compromises their ability to detect problems when they arise, which could potentially lead to patient harm.

While automation bias induced by the presence of CDS increases the risk of patient harm, CDS also has the great potential to reduce errors and improve efficiency. The challenge is to foster appropriate reliance, by relying on CDS when it is correct, disagreeing when it is incorrect, and questioning and verifying when unsure. One specific challenge will be to enable effective and efficient verification of CDS without losing the benefits that correct CDS provides. More research will be needed on how to best assist clinicians with this crucial task. Clinicians should be mindful of the limitations of CDS and the possibility that it can fail. They should be ever-vigilant and ready to verify should uncertainty arise.

6.8 Chapter 6 References

1. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association* 2017;**24**(2):423-31 doi: 10.1093/jamia/ocw105

2. Lyell D, Magrabi F, Coiera E. Reduced Verification of Medication Alerts Increases Prescribing Errors. *Applied Clinical Informatics* 2019;**10**(01):066-76 doi: 10.1055/s-0038-1677009

3. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: An attentional integration. *Human Factors* 2010;**52**(3):381-410 doi: 10.1177/0018720810376055

4. Thackray RI, Touchstone RM. Detection efficiency on an air traffic control monitoring task with and without computer aiding. *Aviation, Space, and Environmental Medicine* 1989;**60**(8):744-48

5. Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced "complacency.". *The International Journal of Aviation Psychology* 1993;**3**(1):1-23 doi: 10.1207/s15327108ijap0301_1

6. Molloy R, Parasuraman R. Monitoring automation failures: effects of automation reliability and task complexity. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 1992 Oct 12-16; Atlanta, GA, USA. Human Factors And Ergonomics Society.

7. Lyell D, Magrabi F, Raban MZ, Pont LG, Baysari MT, Day RO, Coiera E. Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making* 2017;**17**(1):28 doi: 10.1186/s12911-017-0425-5

8. Lyell D, Magrabi F, Coiera E. The effect of cognitive load and task complexity on automation bias in electronic prescribing. *Human Factors* 2018;**60**(7):1008 - 21 doi: 10.1177/0018720818781224

9. Mosier KL, Skitka LJ. Human decision makers and automated decision aids: Made for each other. In: Parasuraman R, Mouloua M, eds. Automation and human performance: Theory and applications. Hillsdale, NJ, England: Lawrence Erlbaum Associates, 1996:201-20.

10. Pelayo S, Kaipio J, Section Editors for the IYSoHF, Organizational I. Findings from the 2018 Yearbook Section on Human Factors and Organizational Issues. *IMIA Yearbook of Medical Informatics 2018* 2018;**27**(01):079-82 doi: 10.1055/s-0038-1667074

11. Sweller J, Ayres P, Kalyuga S. *Cognitive load theory*. New York: Springer, 2011.

12. Fiske ST, Taylor SE. *Social cognition*. New York: Random House, 1984.

13. Gigerenzer G. Why Heuristics Work. *Perspectives on Psychological Science* 2008;**3**(1):20-29 doi: 10.1111/j.1745-6916.2008.00058.x

14. Bagheri N, Jamieson GA. The impact of context-related reliability on automation failure detection and scanning behaviour. IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583); 2004 Oct 10-13.

15. Bagheri N, Jamieson GA. Considering subjective trust and monitoring behavior in assessing automation-induced "complacency". In: Vincenzi DA, Mouloua M, Hancock PA, eds. Human Performance, Situation Awareness and Automation: Current Research and Trends, Vol 2. Mahwah: Lawrence Erlbaum Associates, 2004:54-59.

16. Bahner J, Huper A-D, Manzey D. Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies* 2008;**66**(9):688-99 doi: 10.1016/j.ijhcs.2008.06.001

17. Bahner J, Elepfandt MF, Manzey D. Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2008 Sep 22-26; New York, NY, United states. Human Factors And Ergonomics Society.

18. Manzey D, Reichenbach J, Onnasch L. Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience. *Journal of Cognitive Engineering and Decision Making* 2012;**6**(1):57-87 doi: 10.1177/1555343411433844

19. Reichenbach J, Onnasch L, Manzey D. Misuse of automation: The impact of system experience on complacency and automation bias in interaction with automated aids. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2010 Sep 27 - Oct 1; San Francisco, CA, United states. Human Factors And Ergonomics Society.

20. Reichenbach J, Onnasch L, Manzey D. Human performance consequences of automated decision aids in states of sleep loss. *Human Factors* 2011;**53**(6):717-28 doi: 10.1177/0018720811418222

21. Mack A, Rock I. *Inattentional blindness*. Cambridge, MA: MIT Press, 1998.

22. Mosier KL, Skitka LJ, Heers S, Burdick M. Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology* 1998;**8**(1):47-63 doi: 10.1207/s15327108ijap0801_3

23. Harris WC, Goernert PN. The effect of levels of automation on supervisory performance in a multi-task environment. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 1998 Oct 5-9; Santa Monica, CA, USA. Human Factors And Ergonomics Society.

24. Wolfstadt JI, Gurwitz JH, Field TS, Lee M, Kalkar S, Wu W, Rochon PA. The Effect of Computerized Physician Order Entry with Clinical Decision Support on the Rates of Adverse Drug Events: A Systematic Review. *Journal of General Internal Medicine* 2008;**23**(4):451-58 doi: 10.1007/s11606-008-0504-5

25. Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The Effect of Electronic Prescribing on Medication Errors and Adverse Drug Events: A Systematic Review. *Journal of the American Medical Informatics Association* 2008;**15**(5):585-600 doi: 10.1197/jamia.M2667

26. van Rosse F, Maat B, Rademaker CMA, van Vught AJ, Egberts ACG, Bollen CW. The Effect of Computerized Physician Order Entry on Medication Prescription Errors and Clinical Outcome in Pediatric and Intensive Care: A Systematic Review. *Pediatrics* 2009;**123**(4):1184-90 doi: 10.1542/peds.2008-1494

27. Sweidan M, Williamson M, Reeve JF, Harvey K, O'Neill JA, Schattner P, Snowdon T. Evaluation of features to support safety and quality in general practice clinical software. *BMC Medical Informatics and Decision Making* 2011;**11**(1):1-8 doi: 10.1186/1472-6947-11-27

28. Wright A, Ai A, Ash J, Wiesen JF, Hickman T-TT, Aaron S, McEvoy D, Borkowsky S, Dissanayake PI, Embi P, Galanter W, Harper J, Kassakian SZ, Ramoni R, Schreiber R, Sirajuddin A, Bates DW, Sittig DF. Clinical decision support alert malfunctions: analysis and empirically derived taxonomy. *Journal of the American Medical Informatics Association* 2017;**25**(5):496-506 doi: 10.1093/jamia/ocx106

29. Wright A, Hickman T-TT, McEvoy D, Aaron S, Ai A, Andersen JM, Hussain S, Ramoni R, Fiskio J, Sittig DF, Bates DW. Analysis of clinical decision support system malfunctions: a case series and survey. *Journal of the American Medical Informatics Association* 2016;**23**(6):1068-76 doi: 10.1093/jamia/ocw005

30. Kassakian SZ, Yackel TR, Gorman PN, Dorr DA. Clinical decisions support malfunctions in a commercial electronic health record. *Appl Clin Inform* 2017;**08**(03):910-23 doi: 10.4338/ACI-2017-01-RA-0006

31. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of Drug Safety Alerts in Computerized Physician Order Entry. *Journal of the American Medical Informatics Association* 2006;**13**(2):138-47 doi: 10.1197/jamia.M1809

32. Nanji KC, Slight SP, Seger DL, Cho I, Fiskio JM, Redden LM, Volk LA, Bates DW. Overrides of medication-related clinical decision support alerts in outpatients. *Journal of the American Medical Informatics Association* 2014;**21**(3):487-91 doi: 10.1136/amiajnl-2013-001813

33. Bailey NR, Scerbo MW. Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science* 2007;**8**(4):321-48 doi: 10.1080/14639220500535301

34. Bailey NR, Scerbo MW, Freeman FG, Mikulka PJ, Scott LA. Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors* 2006;**48**(4):693-709 doi: 10.1518/001872006779166280

35. Prinzel LJ, III, Freeman FG, Prinzel HD. Individual Differences in Complacency and Monitoring for Automation Failures. *Individual Differences Research* 2005;**3**(1):27-49

36. Rovira E, McGarry K, Parasuraman R. Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors* 2007;**49**(1):76-87 doi: 10.1518/001872007779598082

37. Singh AL, Tiwari T, Singh IL. Effects of automation reliability and training on automation-induced complacency and perceived mental workload. *Journal of the Indian Academy of Applied Psychology* 2009;**35**(spec iss):9-22

38. Singh IL, Sharma HO, Parasuraman R. Effects of manual training and automation reliability on automation induced complacency in flight simulation task. *Psychological Studies* 2001;**46**(1/2):21-27

39. Singh IL, Singh AL, Saha PK. Monitoring performance and mental workload in an automated system. Proceedings of the International Conference onEngineering Psychology and Cognitive Ergonomics; 2007 Jul 22-27; Beijing, China. Springer Verlag.

40. Mosier KL, Skitka LJ, Dunbar M, McDonnell L. Aircrews and automation bias: The advantages of teamwork? *International Journal of Aviation Psychology* 2001;**11**(1):1-14 doi: 10.1207/s15327108ijap1101_1

41. Skitka LJ, Mosier K, Burdick MD. Accountability and automation bias. *International Journal of Human-Computer Studies* 2000;**52**(4):701-17 doi: 10.1006/ijhc.1999.0349

42. Singh AL, Tiwari T, Singh IL. Performance feedback, mental workload and monitoring efficiency. *Journal of the Indian Academy of Applied Psychology* 2010;**36**(1):151-58

43. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 2012;**19**(1):121-27 doi: 10.1136/amiajnl-2011-000089

44. Povyakalo AA, Alberdi E, Strigini L, Ayton P. How to Discriminate between Computer-Aided and Computer-Hindered Decisions: A Case Study in Mammography. *Medical Decision Making* 2013;**33**(1):98-107 doi: 10.1177/0272989x12465490

45. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical Decision Support Systems Could Be Modified To Reduce 'Alert Fatigue' While Still Minimizing The Risk Of Litigation. *Health Affairs* 2011;**30**(12):2310-17 doi: 10.1377/hlthaff.2010.1111

46. Evans JSBT. In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences* 2003;**7**(10):454-59 doi: https://doi.org/10.1016/j.tics.2003.08.012

47. Gray WD, Fu W-T. Soft constraints in interactive behavior: the case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head*,**. *Cognitive Science* 2004;**28**(3):359-82 doi: 10.1207/s15516709cog2803_3

48. Gray WD, Sims CR, Fu W-T, Schoelles MJ. The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review* 2006;**113**(3):461-82 doi: 10.1037/0033-295X.113.3.461

Appendices

Patient scenarios

The full text of the patient scenario is presented in Appendix A. The interruption tasks are presented in Appendix B.

Instructions to participants

The instructions to participants are presented in Appendix C. The instructional video shown to participants can be viewed at https://youtu.be/Ah8KFfUzlDE. This demonstrated how to use the e-prescribing system, a correct CDS alert and how to verify using the drug references.

Simulated e-prescribing system

The experiment required an e-prescribing system which permitted manipulation of the triggering and content of CDS alerts. It also had to facilitate presentation of the prescribing scenarios. Therefore, a simulated e-prescribing system was developed to fulfil these experimental requirements. It was developed as a web application allowing the experiment to be delivered online.

Appendix D provides an overview of the e-prescribing system.

Appendix A Patient scenarios

The development of the patient scenarios including prescribing errors, the task complexity manipulation and the assessment of the severity of prescribing errors is described in the method section of Article II.

The patient scenarios used in the experiment are presented in the text boxes on the following pages. They include the patient's details, a brief case history and a list of medicines the participants are tasked with prescribing. The highlighting of information elements and the numbering of medicines in these text boxes are to facilitate the presentation of the patient scenarios in this appendix and were not displayed to participants. See Figure D-2 and Figure D-3 in Appendix D for examples of how the scenarios appeared to participants.

This appendix also presents the text of CDS alerts shown to participants in correct CDS conditions, which alerted the prescribing error, and the incorrect CDS conditions which provided a false-positive alert. See Figure D-8 for an example of a correct CDS alert, and Figure D-9 for an example of an incorrect CDS alert as they were shown to participants. An explanation of the prescribing errors is also provided.

No personally identifying information is contained in the patient scenarios.

All patient information was made up for this experiment to present participants with the information they would expect in such patient cases.

Scenario A (Low Complexity)

MRS Dorothea M COLLINS
DOB: 13/03/1930, 84 years **Weight:** 71 kg **Height** 162 cm **Gender:** Female
Allergies: Nil

Mrs. Dorothea M Collins is an 84 year old woman who was admitted to hospital with an acute pulmonary oedema. She has atrial fibrillation which was controlled with Digoxin prior to admission.

Please prescribe the following medications:

A1. Pravastatin sodium 20 mg tablet, 1 tablet, PO, once daily.
A2. Digoxin 250 microgram tablet, 2 tablets, PO, three times a day.
A3. Lisinopril 5mg tablet, 1 tablet, PO, once daily.

Box A-1 Scenario A.

Medications:	3
Conditions:	2
Allergies:	0
Observations:	0
History/Other:	1
Total:	**6**

Table A-1 Number of information elements in scenario A

Prescribing error

A2. Digoxin 250 microgram tablet, 2 tablets, PO, three times a day.

WARNING: High Dose

The entered dose is higher than the recommended maintenance dose range.

Box A-2 Alert displayed for prescribing error. Triggered by the prescription of medicine A2 with correct CDS only.

The elderly patient has atrial fibrillation which was controlled with Digoxin prior to admission. The maintenance dose for an elderly patient is 62.5 to 125 micrograms of Digoxin once daily. The dose requested by the scenario is a loading dose (1,500 mcg over 24 hours), which is too high for the patient presented in the scenario.

Severity: Serious

False-positive

A3. Lisinopril 5mg tablet, 1 tablet, PO, once daily.

WARNING: Medicine Contraindicated

This medication is contraindicated for a condition (Pulmonary Oedema) contained in the patient record.

Box A-3 Alert displayed for false-positive. Triggered by the prescription of medicine A3 with incorrect CDS only.

Lisinopril is not contraindicated in patients with pulmonary oedema.

Scenario B (Low Complexity)

MRS Nancy WEST
DOB: 9/01/1938, 76 years **Weight:** 57 kg **Height** 170 cm **Gender:** Female
Allergies: Nil

Mrs Nancy West is a 76 year old woman. She was admitted with a chest infection. She is also receiving ongoing treatment for systolic heart failure.

Her last blood test indicated she has hyperkalemia.

Please prescribe the following medications:

B1. Carvedilol 6.25 mg tablet, 1 tablet, PO, twice daily.
B2. Augmentin Duo Forte (amoxycillin 875 mg + clavulanic acid 125 mg) tablet, 1 tablet, PO, twice daily.
B3. Spironolactone 25mg tablet, 1 tablet, PO, once daily.

Box A-4 Scenario B

Medications:	3
Conditions:	2
Allergies:	0
Observations:	1
History/Other:	0
Total:	**6**

Table A-2 Number of information elements in scenario B.

Prescribing error

B3. Spironolactone 25mg tablet, 1 tablet, PO, once daily.

WARNING: Medicine Contraindicated

This medicine is contraindicated for a condition (Hyperkalaemia) contained in the patient record.

Box A-5 Alert displayed for prescribing error. Triggered by the prescription of medicine B3 with correct CDS only.

Patient has hyperkalemia which is a contraindication for Spironolactone.

Severity: Serious

False-positive

B2. Augmentin Duo Forte (amoxycillin 875 mg + clavulanic acid 125 mg) tablet, 1 tablet, PO, twice daily.

WARNING: Medicine Contraindicated

This medication is contraindicated for a condition (Systolic Heart Failure) contained in the patient record.

Box A-6 Alert displayed for false-positive. Triggered by the prescription of medicine B2 with incorrect CDS only.

Augmentin Duo Forte is not contraindicated in patients with heart failure.

Scenario C (Low Complexity)

MRS Katie Cadman
DOB: 23/07/1962, 53 years **Weight:** 67 kg **Height** 171 cm **Gender:** Female
Allergies: Nil

Mrs Katie Cadman is a 53 year old woman who was admitted to hospital following a fall and is suffering from pain in her right hip.

She has a history of venous thromboembolism and hyperlipidaemia which are adequately managed.

Please prescribe the following medications:

C1. Warfarin Sodium 2 mg tablet, 1 tablet, PO, once daily.
C2. Atorvastatin 10 mg tablet, 1 tablet, PO, once daily.
C3. Ibuprofen 400 mg tablet, 1 tablet, PO, three times daily.

Box A-7 Scenario C.

Medications:	3
Conditions:	3
Allergies:	0
Observations:	0
History/Other:	0
Total:	**6**

Table A-3 Number of information elements in scenario C.

Prescribing error

C1. Warfarin Sodium 2 mg tablet, 1 tablet, PO, once daily ***and***
C3. Ibuprofen 400 mg tablet, 1 tablet, PO, three times daily.

WARNING: Adverse Drug Interaction

This medication has a listed adverse interaction with another already prescribed medication.

There is an adverse drug interaction for Warfarin and Ibuprofen. This combination should be avoided.

Box A-8 Alert displayed for prescribing error. Triggered by the prescription of medicine C1 and C3 with correct CDS only.

Non-steroidal anti-inflammatory drugs (such as Ibuprofen) increase the risk of gastrointestinal bleeding in patients taking Warfarin. The combination should be avoided, especially as better analgesic options are available.

Severity: Significant

False-positive

C2. Atorvastatin 10 mg tablet, 1 tablet, PO, once daily.

WARNING: Medicine Contraindicated

This medication is contraindicated for a condition (Venous Thromboembolism) contained in the patient record.

Box A-9 Alert displayed for false-positive. Triggered by the prescription of medicine C2 with incorrect CDS only.

Atorvastatin is not contraindicated in patients with venous thromboembolism.

Scenario D (Low Complexity)

MR Thomas Chapman
DOB: 21/05/1971, 43 years **Weight:** 103 kg **Height** 176 cm **Gender:** Male
Allergies: Nil

Mr Thomas Chapman is a 43 year old man who presented in the emergency department with a severe headache and vomiting that have persisted for the last 24 hours.

He suffers from peptic ulcer disease with a history of bleeds.

Please prescribe the following medications:

D1. Pantoprazole 40 mg tablet: enteric, 1 tablet, PO, once daily.
D2. Aspirin 300 mg tablet: effervescent, 3 tablets, PO, every 6 hours.
D3. Metoclopramide Hydrochloride 10 mg tablet, 1 tablet, PO, PRN, every 8 hours, maximum 3 tablets per day.

Box A-10 Scenario D.

Medications:	3
Conditions:	3
Allergies:	0
Observations:	0
History/Other:	0
Total:	**6**

Table A-4 Number of information elements in scenario D.

Prescribing error

D2. Aspirin 300 mg tablet: effervescent, 3 tablets, PO, every 6 hours.

WARNING: Medicine Contraindicated

This medication is contraindicated for a condition (Peptic Ulcer Disease) contained in the patient record.

Box A-11 Alert displayed for prescribing error. Triggered by the prescription of medicine D2 with correct CDS only.

Patient has peptic ulcer disease with a history of bleeds for which aspirin increases the risk of gastrointestinal ulceration. There are better analgesic options available.

Severity: Significant

False-positive

D1. Pantoprazole 40 mg tablet: enteric, 1 tablet, PO, once daily.

WARNING: Medicine Contraindicated

This medication is contraindicated for a condition (Severe Vomiting) contained in the patient record.

Box A-12 Alert displayed for false-positive. Triggered by the prescription of medicine D1 with incorrect CDS only.

Pantoprazole is not contraindicated in patients with severe vomiting.

Scenario E (Low Complexity)

MISS Ava Sunderland
DOB: 03/07/1995, 19 years **Weight:** 48 kg **Height** 173 cm **Gender:** Female
Allergies: Nil

Miss Ava Sunderland is a 19 year old female who presented to the emergency department with severe bloody diarrhoea. She has been passing 8 – 10 stools, containing blood and stringy mucus, on a daily basis over the last 3 weeks.

The results of rectal biopsy and examination are consistent with severe ulcerative colitis.

Please prescribe the following medications:

E1. Mesalazine 500 mg tablet: enteric, 1 tablet, PO, three times daily.
E2. Prednisolone 25 mg tablet, 1 tablet, PO, once daily.
E3. Loperamide Hydrochloride 2 mg capsule, 1 capsule, PO, PRN, every four hours, maximum 8 capsules per day.

Box A-13 Scenario E

Medications:	3
Conditions:	2
Allergies:	0
Observations:	1
History/Other:	0
Total:	**6**

Table A-5 Number of information elements in Scenario E.

Prescribing error

E3. Loperamide Hydrochloride 2 mg capsule, 1 capsule, PO, PRN, every four hours, maximum 8 capsules per day.

WARNING: Medicine Contraindicated

This medication is contraindicated for a condition (Ulcerative Colitis) contained in the patient record.

Box A-14 Alert displayed for prescribing error. Triggered by the prescription of medicine E3 with correct CDS only.

Loperamide is contraindicated in patients with ulcerative colitis which poses a risk of toxic megacolon.

Severity: Serious

False-positive

E1. Mesalazine 500 mg tablet: enteric, 1 tablet, PO, three times daily ***and***
E2. Prednisolone 25 mg tablet, 1 tablet, PO, once daily.

WARNING: Adverse Drug Interaction

This medication has a listed adverse interaction with another already prescribed medication.

There is an adverse drug interaction for Mesalazine and Prednisolone. This combination should be avoided.

Box A-15 Alert displayed for false-positive. Triggered by the prescription of medicine E1 and E2 with incorrect CDS only.

There is no documented adverse drug interaction for Mesalazine and Prednisolone.

Scenario F (Low Complexity)

MR Henry O'Connor
DOB: 05/11/1972, 42 years **Weight:** 84 kg **Height** 181 cm **Gender:** Male
Allergies: Sulfonamide

Mr Henry O'Connor is a 42 year old man who was admitted to the emergency department suffering severe back pain. He has a history of major depression which is currently managed pharmacologically.

Please prescribe the following medications:

F1. Phenelzine 15 mg tablet, 1 tablet, PO, three times daily.
F2. Tramadol Hydrochloride 50mg capsules, 2 capsules, PO, PRN, every six hours, maximum 8 capsules per day.
F3. Ramipril 5 mg tablet, 1 tablet, PO, once daily.

Box A-16 Scenario F

Medications:	3
Conditions:	2
Allergies:	1
Observations:	0
History/Other:	0
Total:	**6**

Table A-6 Number of information elements in scenario F.

Prescribing error

F1. Phenelzine 15 mg tablet, 1 tablet, PO, three times daily ***and***
F2. Tramadol Hydrochloride 50mg capsules, 2 capsules, PO, PRN, every six hours, maximum 8 capsules per day.

WARNING: Adverse Drug Interaction

This medication has a listed adverse interaction with another already prescribed medication.

There is an adverse drug interaction for Phenelzine and Tramadol hydrochloride. This combination should be avoided.

Box A-17 Alert displayed for prescribing error. Triggered by the prescription of medicine F1 and F2 with correct CDS only.

The combination of phenelzine and tramadol is contraindicated due to the possibility of causing serotonin toxicity.

Severity: Serious

False-positive

F3. Ramipril 5 mg tablet, 1 tablet, PO, once daily.

WARNING: Adverse Drug Reaction

This patient has an Allergy or Adverse Drug Reaction recorded for this medication.

Box A-18 Alert displayed for false-positive. Triggered by the prescription of medicine F3 with incorrect CDS only.

The patient is allergic to Sulfonamide. However, Ramipril is not contraindicated for this allergy.

Scenario G (High Complexity)

MRS Beverly Elizabeth WALKER
DOB: 19/03/1942, 72 years **Weight:** 62 kg **Height** 176 cm **Gender:** Female
Allergies: Penicillin

Mrs Beverly Elizabeth Walker is a 72 year old woman who was admitted to hospital in a confused state following a fall. She has since been diagnosed with a urinary tract infection.

The microbiology results show the bacteria as E. Coli (gram –ve) UTI which is sensitive to both Ciprofloxacin and Ampicillin.

She has a history of osteoporosis, osteoarthritis, depression and an allergy to penicillin.

Please prescribe the following medications:

G1. Citalopram 20 mg tablet, 1 tablet, PO, once daily.
G2. Ciprofloxacin 250 mg tablet, 1 tablet, PO, twice daily.
G3. Ramipril 2.5 mg tablet, 1 tablet, PO, once daily.
G4. Paracetamol 500 mg tablet, 2 tablets, PO, four times a day.
G5. Atorvastatin 20 mg tablet, 1 tablet, PO, once daily.
G6. Ural (sodium bicarbonate 1.76 g + citrate sodium 630 mg + citrate 720 mg + tartaric acid 890 mg) sachet, 1 sachet, PO, PRN, three times daily, maximum 3 sachets per day.
G7. Panadeine Forte (Codeine Phosphate with Paracetamol Tablet 30 mg-500 mg) tablet, 2 tablets, PO, PRN, every four hours, maximum 8 tablets per day.
G8. Fosamax Plus 7 0mg (alendronate 70 mg + colecalciferol 70 microgram) tablet, 1 tablet, PO, once a week.

Box A-19 Scenario G

Medications:	8
Conditions:	6
Allergies:	1
Observations:	2
History/Other:	0
Total:	**17**

Table A-7 Number of information elements in scenario G.

Prescribing error

G4. Paracetamol 500 mg tablet, 2 tablets, PO, four times a day ***and***
G7. Panadeine Forte (Codeine Phosphate with Paracetamol Tablet 30 mg-500 mg) tablet, 2 tablets, PO, PRN, every four hours, maximum 8 tablets per day.

WARNING: High Dose / Duplicate Substance

Both Paracetamol and Panadine Forte (Codeinie Phosphate 30mg with Paracetamol 500mg) contain the ingredient Paracetamol. The total Paracetamol entered is higher than the recommended dose range.

Box A-20 Alert displayed for prescribing error. Triggered by the prescription of medicines G4 and G7 with correct CDS only.

Prescribed together, these two prescriptions provide for the combined maximum possible dose of 8 grams of paracetamol per day, double the maximum daily dose of 4 grams.

Severity: Significant

False-positive

G2. Ciprofloxacin 250 mg tablet, 1 tablet, PO, twice daily.

WARNING: Adverse Drug Reaction

This patient has an Allergy or Adverse Drug Reaction recorded for this medication.

Box A-21 Alert displayed for false-positive. Triggered by the prescription of medicine G2 with incorrect CDS only.

The patient is allergic to penicillin. While Ciprofloxacin is an antibiotic, it is not contraindicated for allergy to penicillin.

Scenario H (High Complexity)

MR Arthur Lindsay FOOTE
DOB: 18/09/1940, 74 years **Weight:** 102 kg **Height** 178 cm **Gender:** Male
Allergies: Nil

Mr. Arthur Lindsay Foote is a 74 year old male who presented to hospital with melaena and new onset dizziness. He had been on Naprosyn SR 1 g daily for 2 months for joint pain and early morning swelling affecting his metacarpophalangeal joints. On examination he has evidence of synovitis affecting the small joints of his hands.

Blood results confirm anaemia. He has been transfused with 2 units of packed red cells for ongoing melaena. In addition he has also been diagnosed with a peptic ulcer and new onset rheumatoid arthritis.

Please prescribe the following medications:

H1. Simvastatin 40 mg tablet, 1 tablet, PO, once daily.
H2. Temazepam 10 mg tablet, 1 tablet, PO, once daily.
H3. Perindopril Arginine 5 mg tablet, 1 tablet, PO, once daily.
H4. Paracetamol 500 mg tablet, 2 tablets, PO, PRN, every four hours, maximum 8 tablets per day.
H5. Omeprazole 20 mg capsule, 2 capsules, PO, once daily.
H6. Methotrexate 2.5 mg tablets, 3 tablets, PO, once daily.
H7. Folic acid 500 microgram tablet, 2 tablets, PO, once daily.
H8. Prednisolone 5 mg tablet, 2 tablets, PO, once daily.

Box A-22 Scenario H.

Medications:	8
Conditions:	5
Allergies:	0
Observations:	1
History/Other:	3
Total:	**17**

Table A-8 Number of information elements in scenario H.

Prescribing error

H6. Methotrexate 2.5 mg tablets, 3 tablets, PO, once daily.

WARNING: High Dose

The entered dose is higher than the recommended maintenance dose range

Box A-23 Alert displayed for prescribing error. Triggered by the prescription of medicine H6 with correct CDS only.

Patient has new-onset rheumatoid arthritis. For treatment of rheumatoid arthritis, the loading dose of methotrexate is 7.5mg once weekly. This prescribing error requests the drug be administered daily instead of weekly, which would result in an overdose.

Severity: Potentially lethal

False-positive

H4. Paracetamol 500 mg tablet, 2 tablets, PO, PRN, every four hours, maximum 8 tablets per day.

WARNING: Medicine Contraindicated

This medication is contraindicated for a condition (Peptic Ulcer Disease) contained in the patient record.

Box A-24 Alert displayed for false-positive. Triggered by the prescription of medicine H4 with incorrect CDS only.

Patient has newly diagnosed peptic ulcer disease. However, it is not a contraindication for paracetamol.

Scenario I (High Complexity)

MR Jasper Larnach
DOB: 18/09/1949, 65 years **Weight:** 68 kg **Height** 167 cm **Gender:** Male
Allergies: Opioids

Mr Jasper Larnach is a 65 year old male who was admitted to hospital this morning with severe vomiting and diarrhoea resulting in dehydration and disorientation.

He has a history of Parkinson's disease, osteoarthritis and an allergy to opioids. He also had a myocardial infarction 10 years ago and has been treated for heart failure since.

Please prescribe the following medications:

I1. Paracetamol 500 mg tablets, 2 tablets, PO, four times daily.
I2. Metoprolol Succinate tablet 47.5 mg (controlled release), 1 tablet, PO, once daily.
I3. Levodopa 100 mg + Carbidopa Anhydrous 25 mg tablet, 1 tablet, PO, three times daily.
I4. Entacapone 200 mg tablet, 1 tablet, PO, three times daily.
I5. Ramipril 5 mg tablet, 1 tablet, PO, once daily.
I6. Thiamine Hydrochloride 100 mg tablet, 1 tablet, PO, once daily.
I7. Rosuvastatin 20 mg tablet, 1 tablet, PO, once daily.
I8. Metoclopramide Hydrochloride 10 mg tablet, 1 tablet, PO, PRN, three times daily, maximum 3 tablets per day.

Box A-25 Scenario I.

Medications:	8
Conditions:	7
Allergies:	1
Observations:	0
History/Other:	1
Total:	**17**

Table A-9 Number of information elements in scenario I.

Prescribing error

I8. Metoclopramide Hydrochloride 10 mg tablet, 1 tablet, PO, PRN, three times daily, maximum 3 tablets per day.

WARNING: Medicine Contraindicated

This medication is contraindicated for a condition (Parkinson's Disease) contained in the patient record.

Box A-26 Alert displayed for prescribing error. Triggered by the prescription of medicine I8 with correct CDS only.

The patient has a history of Parkinson's disease for which Metoclopramide is contraindicated as it may cause symptoms to worsen.

Severity: Serious

False-positive

I4. Entacapone 200 mg tablet, 1 tablet, PO, three times daily ***and***
I7. Rosuvastatin 20 mg tablet, 1 tablet, PO, once daily.

WARNING: Adverse Drug Interaction

This medication has a listed adverse interaction with another already prescribed medication.

There is an adverse drug interaction for Entacapone and Rosuvastatin. This combination should be avoided.

Box A-27 Alert displayed for false-positive. Triggered by the prescription of medicines I4 and I7 with incorrect CDS only.

There is no documented adverse drug interaction for Entacapone and Rosuvastatin.

Appendix B Interruption tasks

Interruption task X1

Trevor Chamberlain is currently being treated for a hospital acquired infection with Gentamicin.

The first dose was given 24 hours ago, a second dose has not yet been administered. Please review the patient's creatinine clearance and using the table below determine if a second dose should be given.

Clinical Chemistry		
Sodium (mmol/L)	**136**	135-145
Potassium (mmol/L)	**3.9**	3.5-5.0
Chloride (mmol/L)	**99**	95-107
Bicarbonate (mmol/L)	**25**	24-32
Urea (mmol/L)	**10.1**	3.0-8.0
Creatinine (mmol/L)	**160**	60-110
Creatinine Clearance (mL/min)	**54**	97-137

Creatinine clearance (mL/min)	**Dosing interval**	**Maximum number of empirical doses**
greater than 60	24 hours	3 (at 0, 24 and 48 hours)
40 to 60	36 hours	2 (at 0 and 36 hours)
30 to 40	48 hours	2 (at 0 and 48 hours)
less than 30	give initial dose once, then seek expert advice	

Please review the information above. When should the second dose of Gentamicin be given?

- Give second dose now
- Give second dose in 12 hours
- Give second dose in 24 hours
- Do not give second dose

Correct response: Give second dose in 12 hours

Interruption task X2

Arya Sachs is a 6 year old female who requires analgesia. Refer to the tables below and determine the volume (mL) of paracetamol 120mg/5mL oral liquid for Arya.

Average weight according to age for children	
Age	**Weight (kg)**
6 months	8
1 year	9.6
2 years	12
3 years	14.4
4 years	16
5 years	18.4
6 years	20
8 years	25.6
10 years	32
12 years	40

Paracetamol dose by weight for children	
Weight	**Paracetamol dose (mg)**
9.6	144
12	180
14.4	216
16	240
18.4	276
20	300
25.6	384
32	480
40	600
51.2	768

Paracetamol 120mg/5mL conversion table: Dose (mg) to Volume (mL)	
Dose (mg)	**Volume (mL)**
108	4.5
120	5
144	6
180	7.5
216	9
240	10
276	11.5
300	12.5
384	16

Please review the information above. What volume (mL) of paracetamol 120 mg/5 mL oral liquid should be administered?

- 5 mL
- 7.5 mL
- 9 mL
- 10 mL
- 11.5 mL
- 12.5 mL

Correct response: 12.5 mL

Interruption task X3

Harold O'Brien is about to commence Vancomycin for meningitis. Please refer to the test Creatinine Clearance in the test results provided and select the appropriate dose and frequency of administration using the dosing tables below.

Clinical Chemistry		
Sodium (mmol/L)	**140**	135-145
Potassium (mmol/L)	**4.5**	3.5-5.0
Chloride (mmol/L)	**106**	95-107
Bicarbonate (mmol/L)	**28**	24-32
Urea (mmol/L)	**9.2**	3.0-8.0
Creatinine (mmol/L)	**141**	60-110
Creatinine Clearance (mL/min)	**64**	97-137

Creatinine clearance (mL/min)	**Starting maintenance dosage**
more than 90	1.5 g
90 or less	1 g

Creatinine clearance (mL/min)	**Frequency**
more than 60	12-hourly
20 to less than 60	24-hourly
less than 20	48-hourly

Please review the information above. What *dose* of Vancomycin should be given and how *frequently* should it be administered?

- 1.5 g 12-hourly
- 1.5 g 24-hourly
- 1.5 g 48-hourly
- 1 g 12-hourly
- 1 g 24-hourly
- 1 g 48-hourly

Correct response: 1 g 12-hourly

Appendix C Instructions to participants

Instructions

The instructions were presented to participants after they provided informed consent and completed the pre-experiment questionnaire. The written instructions are shown in Figure C-1.

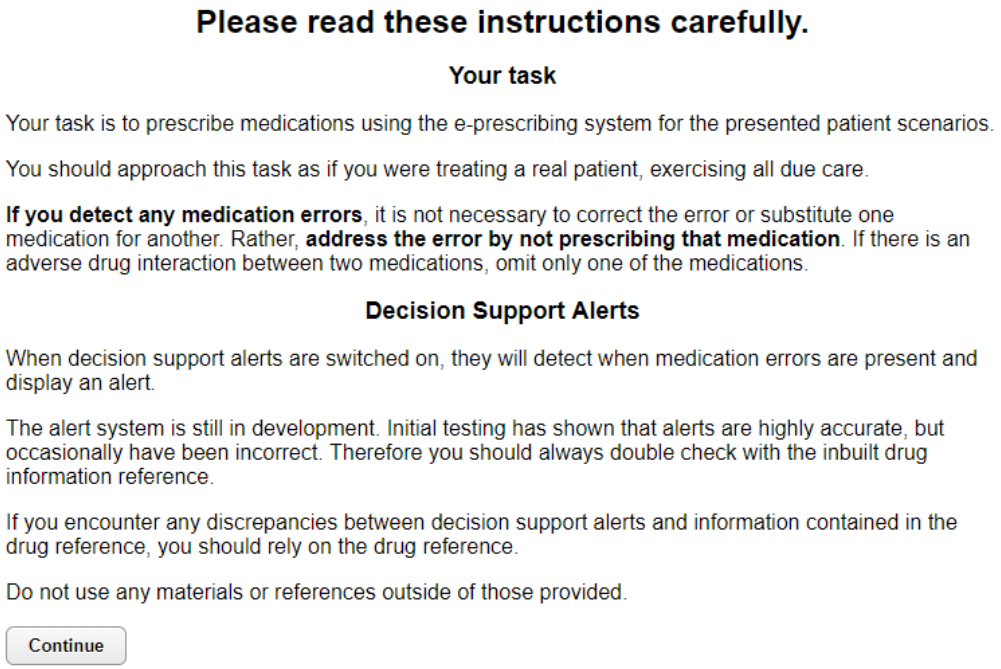

Figure C-1 Screen capture of participant instructions.

These instructions emphasised two points important to ensuring consistency with prior automation bias studies: (1) that CDS could be incorrect, and (2) the method to verify CDS with a non-automated and completely accurate information source,[1, 2] in this experiment the drug reference.

The instructions set out how participants should respond to prescribing errors by not prescribing the affected medicine. Several prescribing errors and CDS false-positives involved an adverse-drug interaction between two medicines; in this event, participants were instructed only to omit one.

The reason for having participants omit medicines containing error was that key experiment dependent variables were focused on the information seeking and processing associated with the task of detecting prescribing errors. Correcting prescribing errors creates another task, which would involve further information seeking and processing as participants search for and evaluate alternative treatments which provide the same therapeutic outcomes, but avoid the contraindication identified.

Video demonstration

After viewing the written instructions, participants were shown a 2-minute and 57-second instructional video on how to use the e-prescribing system. This was presented as an embedded YouTube video to ensure the greatest possible compatibility across different web browsers.

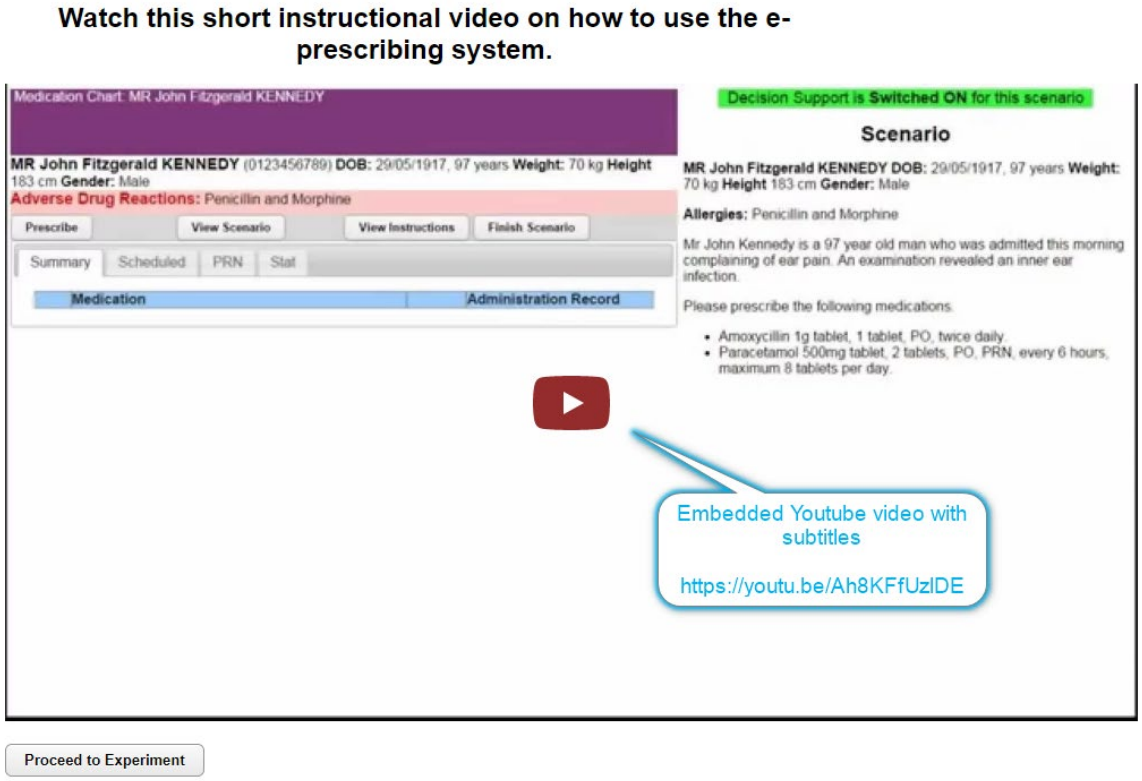

Figure C-2 Screen capture of the instructional video shown after participants read the written instructions.

This video demonstrated:

- How to identify whether CDS alerts are switched on (experimental *CDS assisted* conditions) or off (control conditions).
- An overview of the patient scenarios and the participants' task.
- How to order a prescription using the e-prescribing system.
- How to view the drug reference, including adverse drug interactions.
- How to remove or cease a prescription on the medication chart.
- The functioning of CDS alerts. This also demonstrated:
    - A prescribing error contained within a patient scenario, where Amoxicillin (an antibiotic) was requested for a patient with an allergy to penicillin.
    - Identifying the contraindication using the drug reference.
    - How to resolve CDS alerts.

This video can be accessed on YouTube at the following URL: https://youtu.be/Ah8KFfUzlDE

Orienting participants to CDS assisted and control conditions

The experimental control were conditions in which participants prescribed with no CDS assistance. It was important to clearly differentiate between conditions with and without CDS. This was especially important in relation to the absence of alerts, which, in experimental conditions, indicates that CDS did not detect any prescribing errors, while, in control conditions, there was no CDS checking.

To orient participants to whether CDS was assisting them in the current scenario: (1) all control scenarios (no CDS) were presented together in a block of three, and all experimental scenarios (correct and incorrect CDS) were presented together in a block of six. (2) At the beginning of each block, participants were shown a message explaining whether CDS would be switched on or off.

Figure C-3 depicts the message shown to participants at the beginning of the control scenario block. Figure C-4 is the message shown to participants at the beginning of the experimental block for scenarios assisted by correct or incorrect CDS.

Participants were blinded as to whether the CDS assistance they received was correct or incorrect.

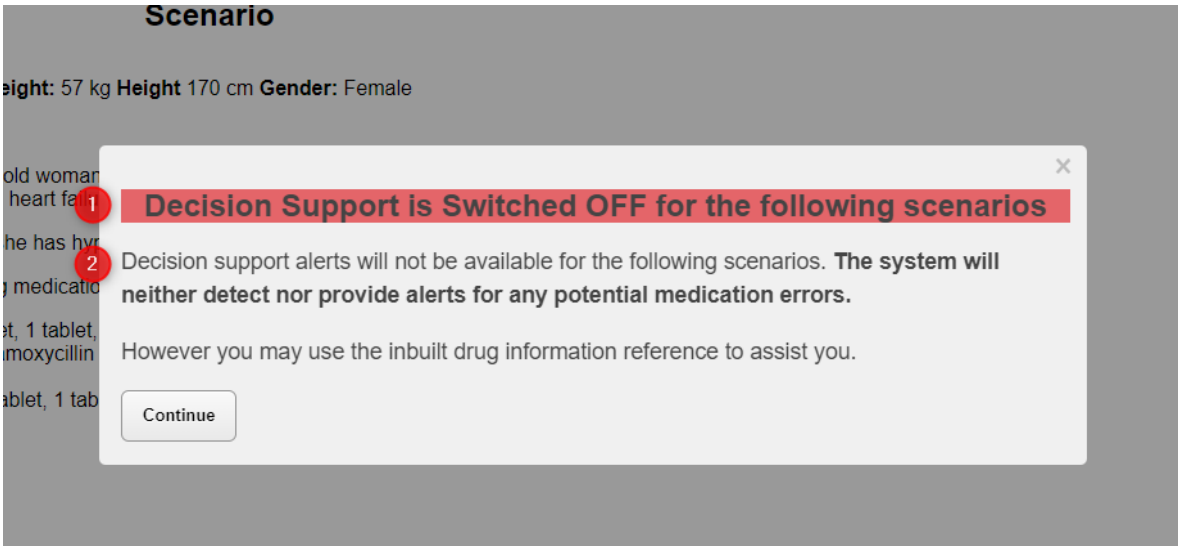

Figure C-3 Screen capture of the CDS status message shown at the beginning of the control (No CDS) block.

1 *CDS status indicator, showing that CDS is switched off.*

2 *The implication of CDS being switched off is explained to participants. The system will not check for, and alert, prescribing errors and they should refer to the provided drug reference.*

The message was displayed as a modal window and had to be acknowledged before the participant could continue.

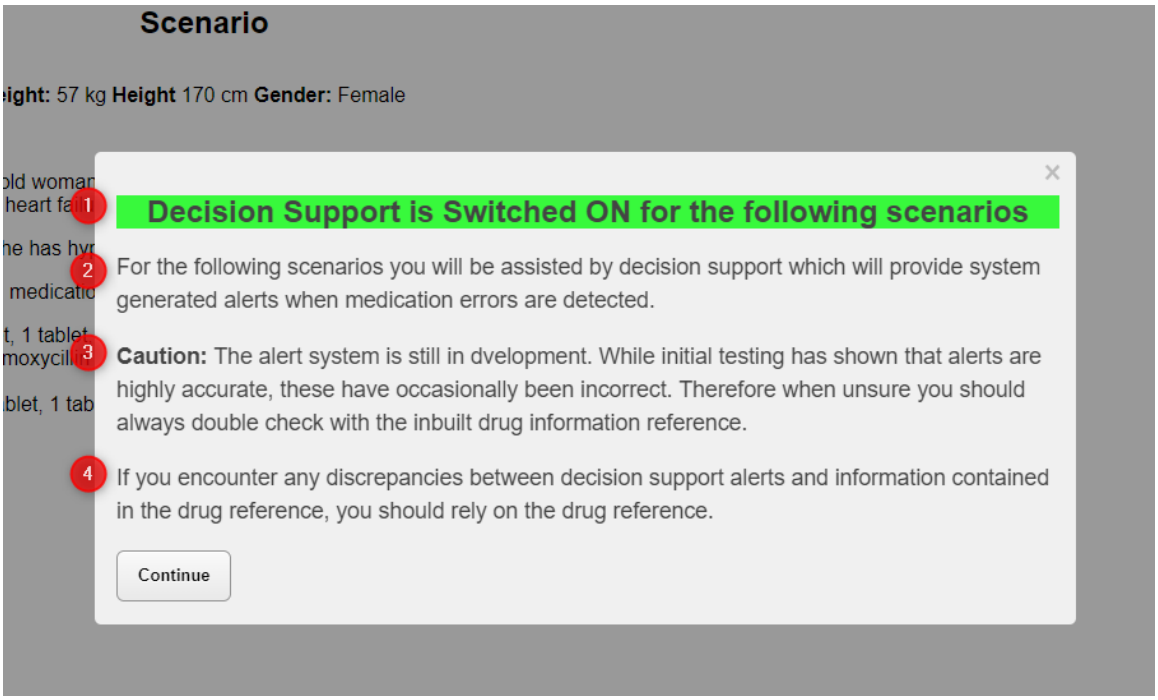

Figure C-4 Screen capture of the CDS status message shown at the beginning of the experimental (correct and incorrect CDS) block.

1 *CDS status indicator, showing that CDS is switched on.*

2 *The significance of this is that CDS will check for, and alert, detected prescribing errors.*

3 *Participants were cautioned that CDS could be incorrect, but that it can be checked using the provided drug references.*

4 *Participants are instructed that they should rely on the drug reference over CDS, whenever there is a conflict between the two.*

List of references

1. Skitka LJ, Mosier K, Burdick MD. Accountability and automation bias. *International Journal of Human-Computer Studies* 2000;**52**(4):701-17 doi: 10.1006/ijhc.1999.0349

2. Bahner J, Huper A-D, Manzey D. Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies* 2008;**66**(9):688-99 doi: 10.1016/j.ijhcs.2008.06.001

Appendix D Overview of the simulated e-prescribing system

A key requirement of the experiment was the ability to manipulate the triggering and content of CDS alerts within the e-prescribing system and facilitate the presentation of the patient scenarios.

A simulated e-prescribing system was developed to satisfy these requirements. Functionality was limited to only those needed by the experiment, in particular, those functions needed to prescribe and cease medicines, view relevant drug references and processes for working with CDS alerts. Functionality not required by the experiment was omitted, for example, the medication administration record.

The experiment was presented to participants as the evaluation of an e-prescribing system currently in development; this formed part of the pretext for the caution that CDS had failed in testing which was described in the instructions.

Figure D-1 presents a flow diagram of user interactions with the simulated e-prescribing system as they performed the experimental task.

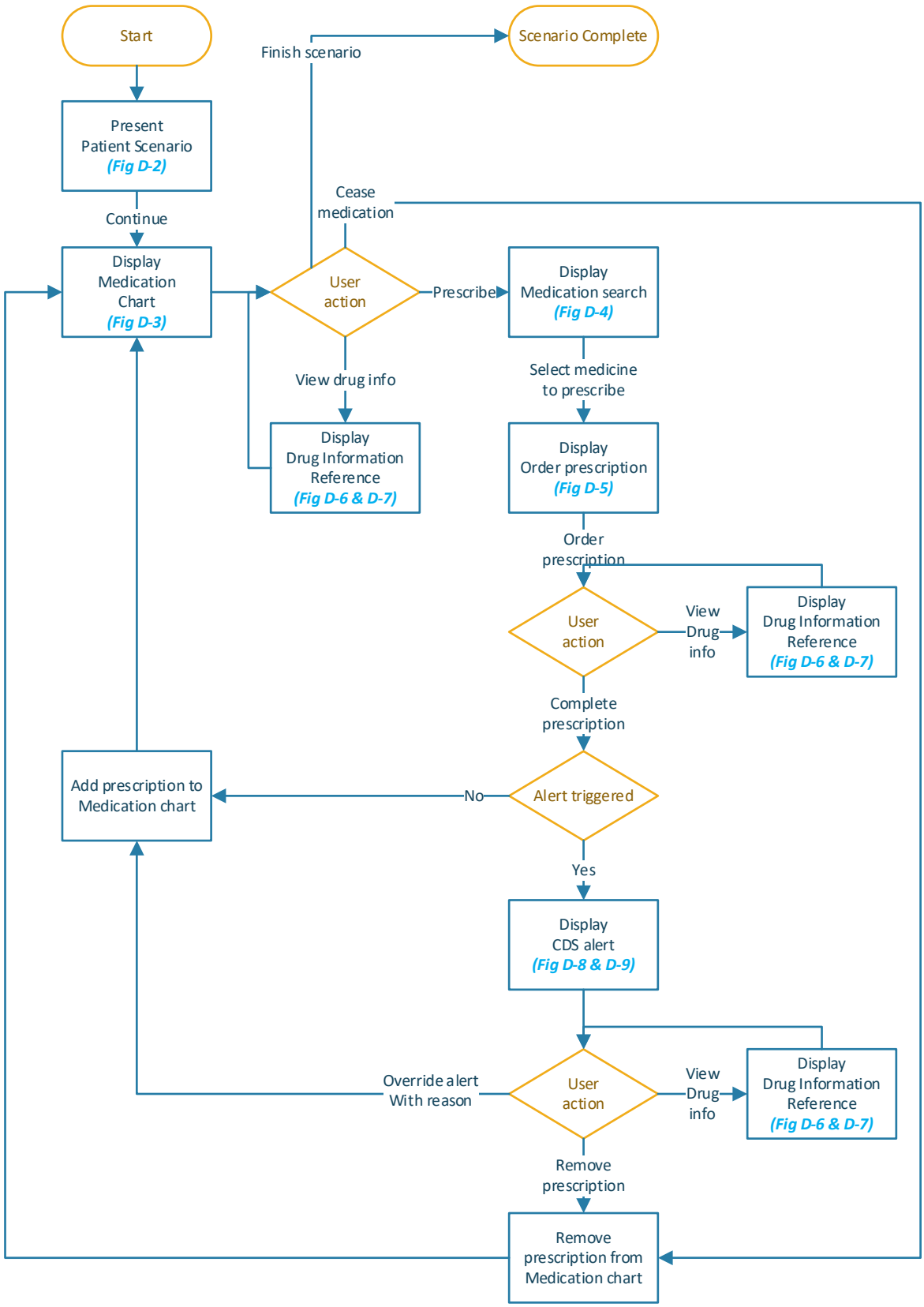

Figure D-1 Flow diagram of the e-prescribing system.

Patient Scenario

Each scenario (or trial) began with the presentation of the patient scenario (Figure D-2) in which the participant was tasked with prescribing medications.

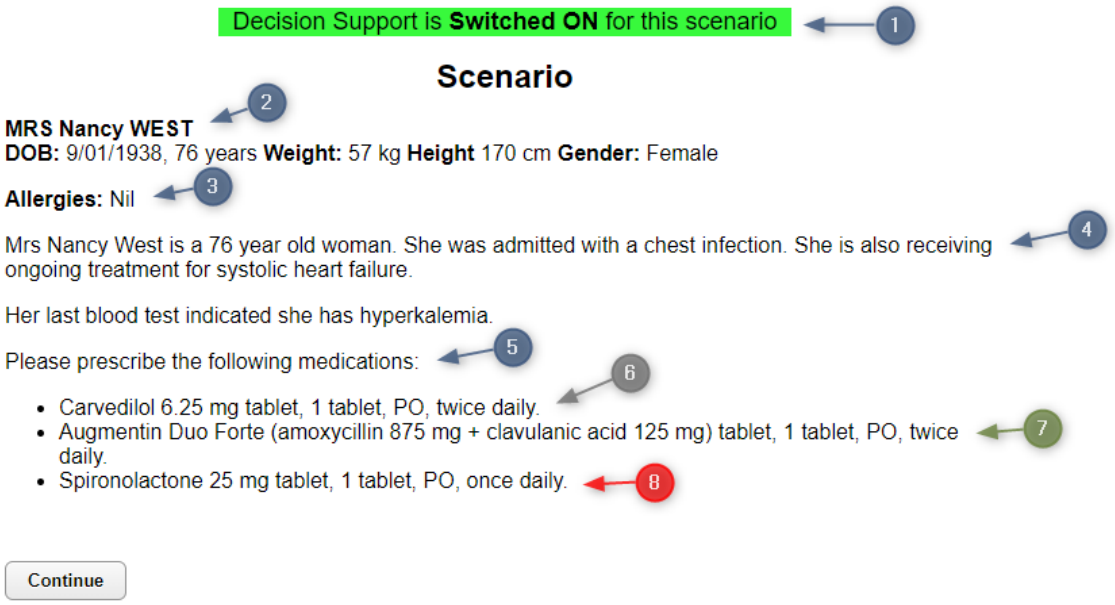

Figure D-2 Screen capture of the initial presentation of the patient scenario.
This figure depicts the presentation of Scenario B (see Appendix A.)

1 *The CDS status indicator.*

2 *Patient details.*

3 *Medicines to which the patient is allergic or to which they have experienced an adverse reaction.*

4 *Details of the case, including patient history, the reason for admission and conditions to be treated.*

5 *A list of medicines participants are tasked with prescribing.*

6 *For this scenario (Scenario B) this medicine is unaffected by prescribing error.*

7 *For this scenario (Scenario B), Augmentin Duo Forte is the designated 'false-positive' medicine. In the incorrect CDS condition, its prescription will trigger a false-positive alert warning that it is contraindicated in patients with systolic heart failure.*
This is not a true contraindication and is not supported by the drug reference.

8 *For this scenario (Scenario B), Spironolactone is the prescribing error. It is contraindicated in patients with hyperkalaemia. In the correct CDS condition, an alert is displayed with this information. The drug reference documents this contraindication, providing support for the alert.*

Medication chart

After viewing the patient scenario, participants started their task on the medication chart screen (Figure D-3). The scenario starts with an empty medication chart. From this screen, they can initiate the prescription of new medications, view the list of prescribed medications, as well as view the drug reference or cease prescribed medicines.

The medication chart screen (Figure D-3) and prescribing screens (Figure D-4 and Figure D-5) were presented side-by-side with the patient scenario. This was to eliminate the potential for split-attention effects, whereby people have to integrate information presented on separate screens in working memory, which increases cognitive load.[1]

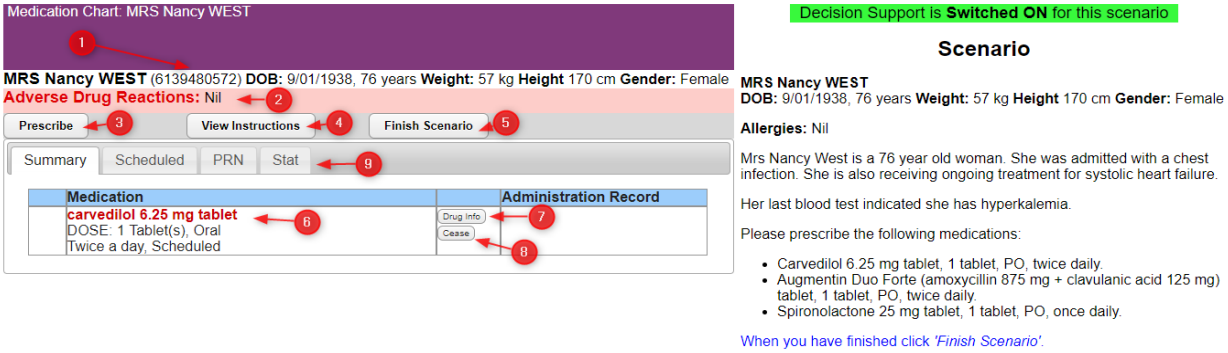

Figure D-3 Screen capture of the e-prescribing system medication chart.

1 *Patient details.*

2 *List of medicines to which the patient is allergic or has had an adverse reaction to.*

3 *The 'Prescribe' button initiates the process to prescribe a medicine for the current patient. This takes the participant to the 'Prescribe, search for medication' screen (Figure D-4).*

4 *The 'View instructions' button opens a modal window allowing the participant to view the instructions and instructional video again if they require (Figure C-1 and Figure C-2).*

5 *Participants indicate they have completed the current scenario by pressing the 'Finish Scenario' button. After confirming the action, they complete the Cognitive Load Inventory, following which the next scenario is initiated.*

6 *Prescriptions which have been ordered, providing the generic drug name, strength and form, dose, route and frequency of administration.*

7 *The 'Drug info' button opens the drug reference for the current medicine.*

8 *The 'Cease' button ceases or removes the current medicine.*

9 *Participants can filter ordered medicines by whether they are regularly scheduled, PRN (or as needed) or stat (to be administered immediately). By default, the summary tab is shown which displays all orders.*

Prescribe: medication search

When the participant initiates a new prescription, they are first taken to the medication search and selection screen (Figure D-4). On this screen, the participant searches for, and selects, the medicine they wish to prescribe.

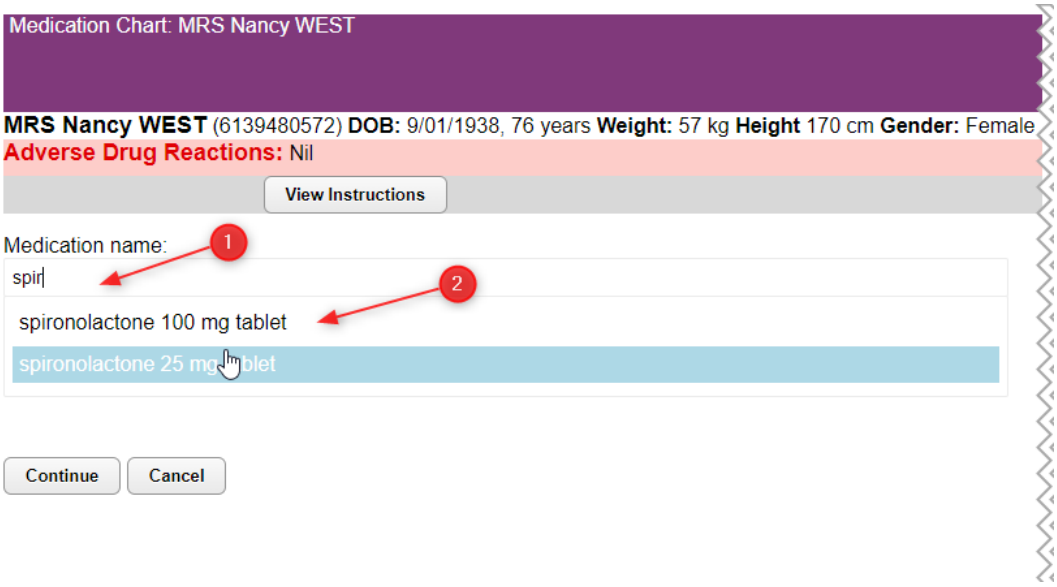

Figure D-4 Screen capture of the medication search screen.
Here the participant searches for, and selects, the medicine they wish to prescribe.

1 *Search for medicine name.*

2 *Search results appear as the participant types in the medicine name. The search results provide all possible forms and strengths of the matching medicine which are available.*

As with the medication chart, the patient scenario is displayed to the right of the search for medicine screen but is not shown in this figure.

Prescribe: order screen

Once the participant has selected the medication they wish to prescribe, they are directed to the order screen (Figure D-5) where they specify the details of the order, including the prescription type, dose, frequency and route.

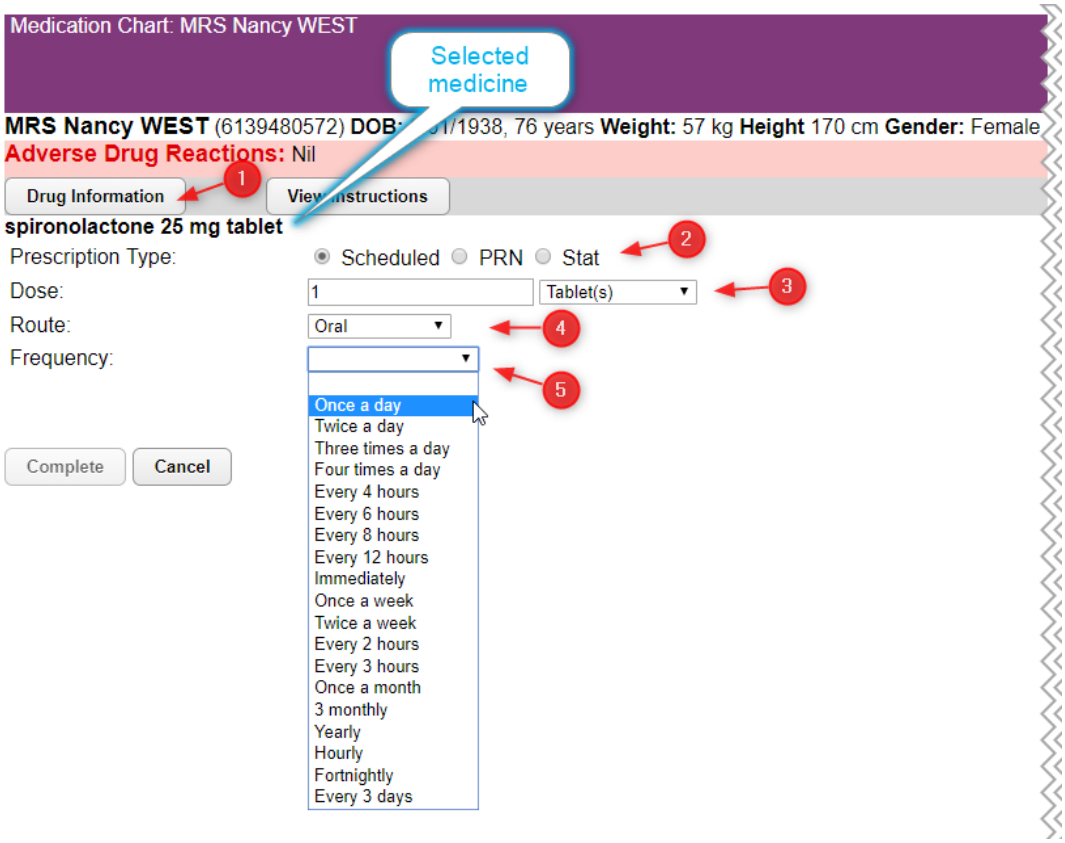

Figure D-5 Screen capture of the medication order screen.
Here, the participants order the prescription of the medicine they selected in the previous step.

1 *Opens the drug reference for the currently selected medicine.*

2 *Prescription type: whether the medicine is administered as scheduled, as needed (PRN) or as a single dose given immediately (stat).*

3 *Dose: expressed as the number of the select units, in this example, 1 tablet. The units field is pre-populated based on the currently selected medicine.*

4 *Route: The route of administration, this field is also pre-populated based on the currently selected medicine.*

5 *Frequency: The frequency of administration.*

If the prescription type is PRN, an additional field appears where the participant must specify the maximum dose per day.

The patient scenario is displayed to the right of the medication order screen but is not shown in the current figure.

Drug reference viewer

The drug reference viewer displays monographs from the *Australian Medicines Handbook*,[2] for the currently selected medicine. The *Australian Medicines Handbook* references the Australian formulary, is evidence-based and is widely utilised in Australian clinical practice.[3] This provided participants with sufficient information to verify prescription safety independent of CDS and to verify the correctness of CDS alerts.

The references were checked to ensure they clearly identified all prescribing errors and contradicted all false-positive alerts in the patient scenarios (see Figure D-6 and Figure D-7).

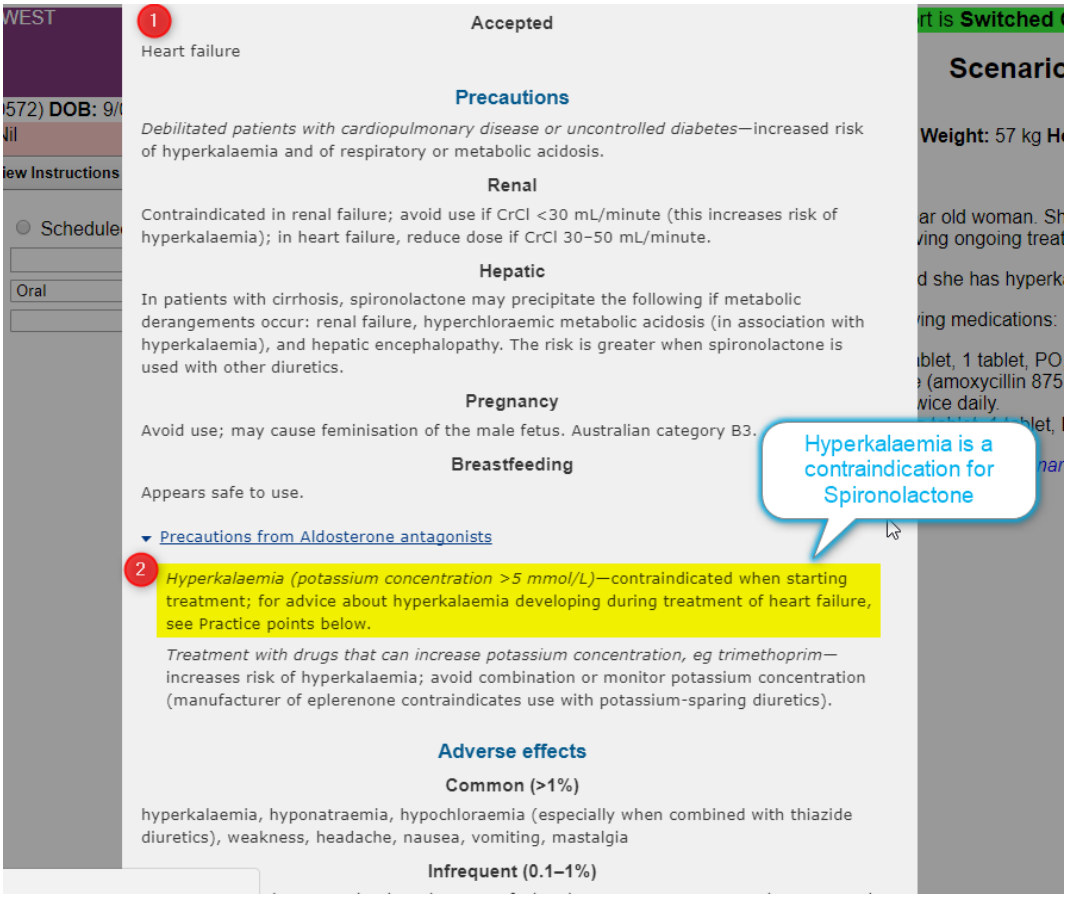

Figure D-6 Screen capture of the drug reference viewer, displaying the reference for spironolactone, the prescribing error in Scenario B.

The drug reference is displayed as a modal window over the top of the e-prescribing system.

❶ *The drug monographs presented in the drug reference are from the Australian Medicines Handbook.[2]*

❷ *The prescribing errors inserted into each scenario were identifiable with information contained in the drug reference. (The relevant section is highlighted for presentation here but was not highlighted for participants.)*

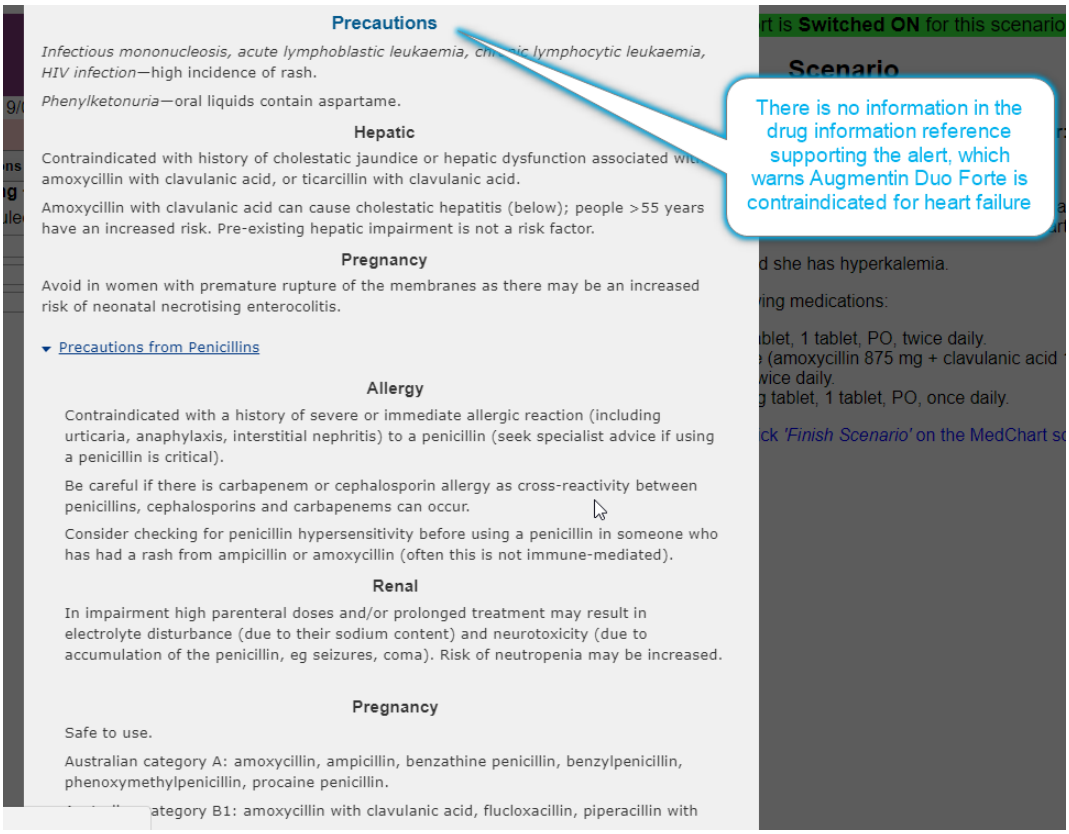

Figure D-7 Screen capture of the drug reference viewer, displaying the reference for Augmentin Duo Forte.

Augmentin Duo Forte receives a false-positive alert in Scenario B when assisted by incorrect CDS which provided a warning indicating the presence of a prescribing error when there was none. The reason displayed by the alert was made up and is clinically incorrect. Consequently, the alert was not supported by the drug reference. In the context of this patient scenario, Augmentin Duo Forte was a necessary medicine.

CDS alerts

Alerts were triggered by the experimental condition and the medicine prescribed. Correct CDS triggered alerts when the medication containing the prescribing error was prescribed, while incorrect CDS triggered alerts when the designated false-positive medication was ordered (see Article II).

Figure D-8 provides an example of an alert triggered by correct CDS in Scenario B, while Figure D-9 shows an alert triggered by incorrect CDS in the same scenario. Participants were blinded to whether CDS was correct or incorrect. Both true- and false-positive alerts were presented identically; the only way to differentiate between them was to verify the content for the alert using the drug reference.

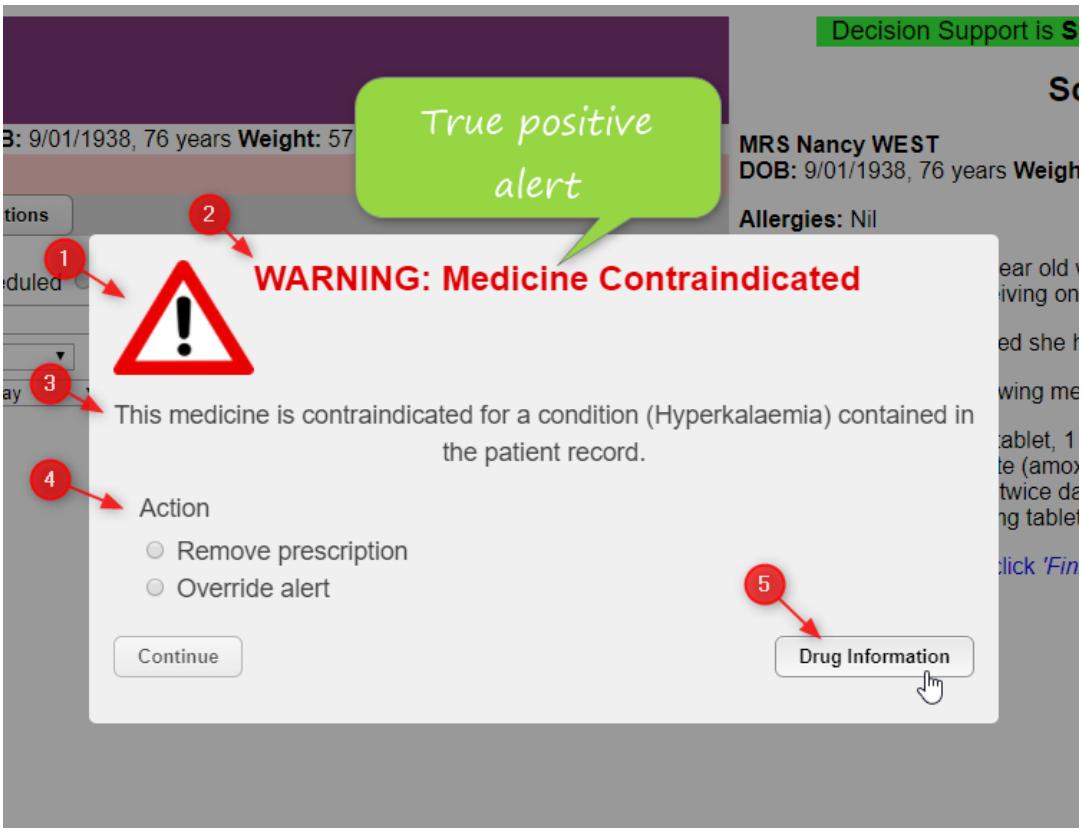

Figure D-8 Screen capture of a CDS alert triggered by a prescribing error with correct CDS.

The alerts were interruptive, displayed in a modal window over the top of the e-prescribing system and had to be resolved before the participant was allowed to close it and continue.

This figure shows a true-positive alert from correct CDS, which correctly identifies the prescribing error in Scenario B, where Spironolactone is contraindicated because the patient had a test result showing hyperkalaemia.

1 *Alert icon.*

2 *Nature of the alert.*

3 *Specific reason for the alert.*

4 *Actions to resolve the alert. The participant could choose to remove the prescription or override the alert and prescribe the medicine.*

5 *The alert provided direct access to the relevant drug reference.*

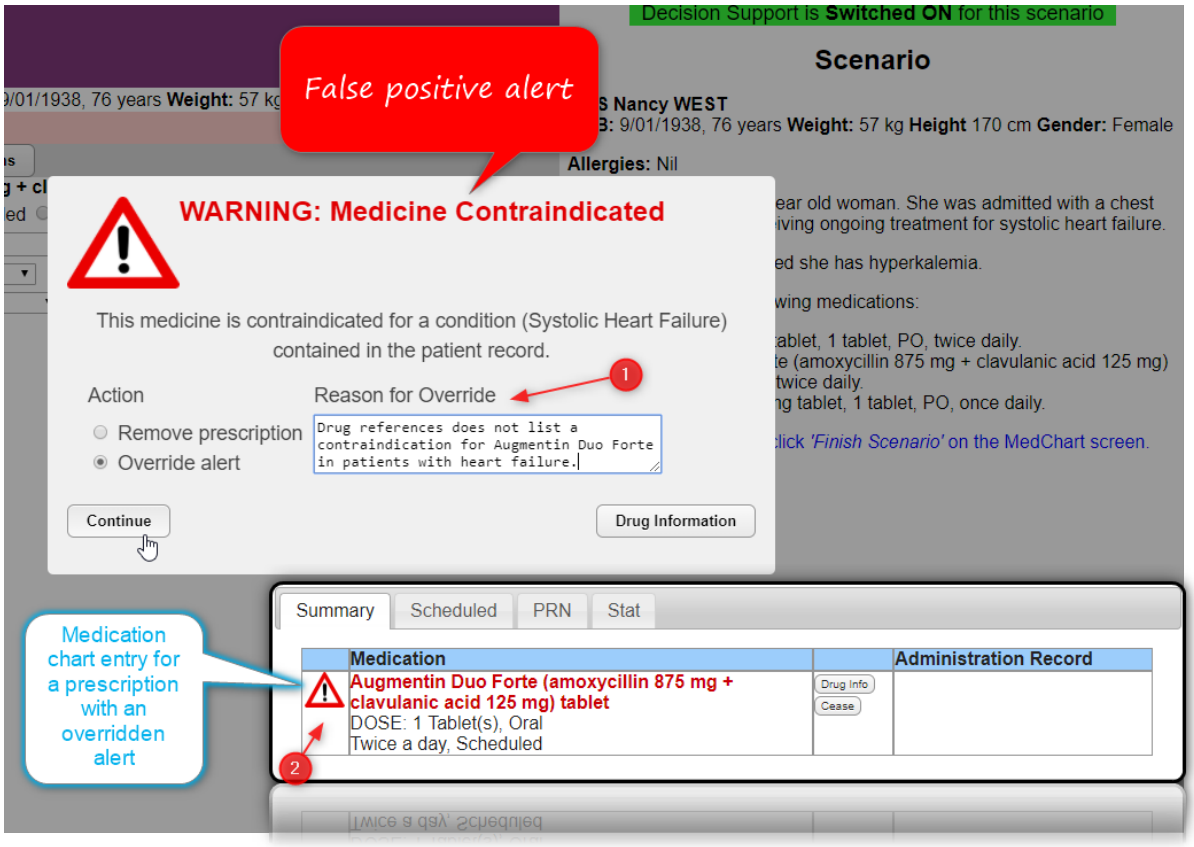

Figure D-9 Screen capture of a CDS alert triggered by the designated false-positive with incorrect CDS.

This example is a false-positive alert from incorrect CDS. Augmentin Duo Forte is unaffected by prescribing errors for the patient in Scenario B. The reason for the alert is false and is contradicted by the drug reference.

1 *Resolving an alert by overriding it requires the participant to provide a reason for overriding. Once overridden, the prescription is added to the medication chart.*

2 *The alert icon indicates that CDS detected an error in this order. Clicking the icon displays the details of the alert without the resolution actions to remove the prescription or override the alert. If the participant wanted to remove the order, they could do so from the medication chart using the 'cease' button.*

List of References

1. Sweller J, Ayres P, Kalyuga S. *Cognitive load theory*. New York: Springer, 2011.

2. Australian Medicines Handbook Pty Ltd. Australian Medicines Handbook 2015 (online). Secondary Australian Medicines Handbook 2015 (online) 2015 January. Retrieved from http://amhonline.amh.net.au/.

3. Day RO, Snowden L. Where to find information about drugs. *Australian Prescriber* 2016;**39**(3):88-95 doi: 10.18773/austprescr.2016.023

Appendix E Human research ethical approvals

The experimental research presented in this thesis complied with the National Statement on Ethical Conduct in Human Research,[1] and was approved by the Macquarie University Human Research Ethics Committee (Ref: 5201401029) and the University of New South Wales Medical and Community Human Research Ethics Advisory Panel (Ref: 2014-7-32). Copies of the approval letters are presented in this appendix.

1. National Health and Medical Research Council, Australian Research Council, Australian Vice-Chancellors' Committee. National Statement on Ethical Conduct in Human Research. *Council NHaMR* 2007 (Updated 2015) www.nhmrc.gov.au/guidelines/publications/e72

Appendix E of this thesis has been removed as it may contain sensitive/confidential content