The study of clinical work processes in hospitals

Methods and applications of the quantitative observational approach

Scott R. Walter

B.A., Grad. Dip. Ed., M.Biostat.

Centre for Health Systems and Safety Research
Australian Institute of Health Innovation
Faculty of Medicine and Health Sciences
Macquarie University

July 2016

This work is submitted for the degree of Doctor of Philosophy

Contents

Abstract

Clinicians work in complex, dynamic work environments where synchronous communication and the management of competing time-constrained demands in a team environment are fundamental to safe healthcare delivery. Previous empirical research on clinical workflow has investigated the characteristics of clinical tasks and the context in which they occur as potential contributors to the risk of error. Strategies adopted by clinicians, such as the use of interrupting and multitasking, have been a particular focus of such study given their prevalence in clinical work and widespread perception of their negative impact on efficiency and safety. However, our understanding of the role and effects of these behaviours in complex healthcare settings is far from complete. A major impediment to progressing knowledge has been a lack of quantitative observational methodology, namely the design of studies using direct observations that adequately encompass the complexity of clinical work, and the use of appropriate data analytic techniques. This thesis aimed to advance direct observational methods and related statistical analysis techniques relevant to work processes in non-experimental settings, and to apply those methods to the study of everyday clinical practices.

Drawing upon an extensive examination of cross-disciplinary research, a clinically-relevant conceptualisation of the work process was proposed that unifies several concepts previously used to study interruptions and multitasking in healthcare. Second, this framework was applied to analyse existing datasets comprising over one thousand hours of observations of clinicians, yielding fresh insights about their work practices. Third, an observational study of doctors was designed around the conceptualisation and conducted in the emergency department of a tertiary hospital. This extended previous approaches to provide a more comprehensive analysis of strategies used by clinicians in response to a range of disruptive events. Fourth, an overarching theme of this thesis is to identify ways in which existing statistical methodology can be both better applied and extended to expand the scope of research questions that can be tackled on work processes in healthcare. The analyses

above implemented statistical modelling techniques in a way not previously used to study clinical work. Furthermore, a new technique was developed to assess the impact of disruptive events on task completion time, and was then applied to data from observations of doctors in a range of hospital settings.

Through the application of improved observational and statistical methods this thesis has progressed debates about the conceptualisation of clinical workflow, identified factors that influence clinicians' strategies to manage disruptive events in a range of healthcare settings, and better quantified the impact of interruptions on task completion time. This has provided a more sophisticated understanding of the relationships between work behaviours, work efficiency and error production. Moreover, the methodological progress enables future creation of knowledge necessary for safety improvement.

Declaration

I certify that the content of this thesis has not previously been submitted for a higher degree to any other university or institution.

I also declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at Maqcquarie University or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work.

The research presented in chapter 5 was approved by the South East Sydney Local Health District Human Research Ethics Committee on 21^{st} May 2014 (Ref. 13/310 [HREC/13/POWH/674]).

Scott R. Walter

Co-author agreement

The substantive part of this thesis comprises five publications, each of which was prepared in collaboration with several co-authors. The doctoral candidate is the first author on all publications and made the largest contribution in each case. Further details of co-author contributions are described in the introductory sections preceding each paper. Each coauthor has given permission to include these publications in this thesis as indicated by the signatures on the following page.

All co-authors agree to the inclusion in this doctoral thesis of publications to which they contributed.

Co-author name	**Signature**	**Date**
Johanna I Westbrook	J Westbrook	16th May 2016
William TM Dunsmuir	[illegible]	18th May 2016
Ling Li	Ling Li	18th May 2016
Bruce M Brown	B M Brown	17th June 2016
Magdalena Z Raban	[illegible]	24th May 2016
Heather E Douglas	H Douglas	19th May 2016

Acknowledgements

This work represents my figurative academic progress from, say, point n to point $n+1$. So as a foreground to acknowledging individuals who supported that progression, it is important to first recognise that getting to point n was attributable more to circumstance and the support of others than to my own efforts or abilities.

Naturally the first people to acknowledge are my supervisors: Johanna Westbrook and William Dunsmuir. Throughout the candidacy they have been nothing but supportive. Despite their extensive professorial duties, they were always available for supervision meetings, responded promptly and thoughtfully to drafts, and struck a well-judged balance between management and autonomy. As a team they formed the perfect supervisory complement, both in their domains of knowledge and the nature of their input into my work. William has an acute ability to cut to the essence of a statistical problem and, thanks to his patience with my patchy statistical knowledge, I now have robust understanding of the process of developing statistical methodology and an intensified motivation to continue in that vein. In addition to her expert domain-specific guidance, Johanna also imparted invaluable lessons about research communication including, but not limited to, the art of distilling out the key messages of a study and how to maintain a wide-angle perspective on the relevance and direction of my research.

I would also like to thank my co-authors - our collaborative combinations were more than the sum of their parts. Thanks also to the people in the Australian Institute of Health Innovation, especially those in the Centre for Health Systems and Safety Research. The value of the supportive and collaborative work environment that they create is without measure.

Finally, no words can possibly express the full extent of the support, encouragement and inspiration of my family, other than to say a heartfelt and immeasurable *thank you*.

List of abbreviations

AE	Adverse event
CI	Confidence interval
ED	Emergency department
HCI	Human-computer interaction
OR	Odds ratio
RMO	Resident medical officer
RN	Registered nurse
SRMO	Senior resident medical officer
WOMBAT	Work Observation Method By Activity Timing

List of outputs

Articles included in this thesis

Walter SR, Dunsmuir WTM, Westbrook JI. 2015. Studying interruptions and multitasking in situ: The untapped potential of quantitative observational studies. *Int J Hum Comput St*, 79: 118-125.

Walter SR, Li L, Dunsmuir WTM, Westbrook JI. 2014. Managing competing demands through task-switching and multitasking: A multi-setting observational study of 200 clinicians over 1000 hours. *BMJ Qual Saf*, 23: 231-241.

Walter SR, Raban MZ, Dunsmuir WTM, Douglas HE, Westbrook JI. 2017. Emergency doctors' strategies to manage competing workload demands in an interruptive environment: an observational workflow time study. *Appl Ergon*, 58: 454-460.

Walter SR, Brown BM, Dunsmuir WTM. 2016. Assessing the impact of task-switching on completion of clinical tasks in the presence of length bias. *J Roy Stat Soc C*, submitted.

Walter SR, Dunsmuir WTM, Westbrook JI. 2016. Assessing the effect of interruptive events on task completion time: a multi-site study. Proceedings of the International Conference on Healthcare Systems Ergonomics and Patient Safety (HEPS 2016), Toulouse, France, 5-7 October: p. 364-371.

Conference presentations related to this thesis

Walter SR, Dunsmuir WTM. 2014. Assessing the impact of task-switching on task length in the presence of length bias. Oral presentation at the Australian Statistical Conference in conjunction with the Institute of Mathematical Statistics Annual Meeting (ASC-IMS 2014), Sydney, Australia, 7-10 July.

Walter SR, Dunsmuir, WTM, Westbrook JI. 2016. Assessing the effect of interruptive events on task completion time: a multi-site study. Oral presentation at the International Conference on Healthcare Systems Ergonomics and Patient Safety (HEPS 2016), Toulouse, France, 5-7 October.

Other articles relevant to this work

Raban MZ, **Walter SR**, Douglas HE, Strumpman D, Mackenzie J, Westbrook JI. 2015. Measuring the relationship between interruptions, multitasking and prescribing errors in an emergency department: A study protocol. *BMJ open*, 5(10): e009076.

Richardson LC, Lehnbom EC, Baysari MT, **Walter SR**, Day RO, Westbrook JI. 2016. A time and motion study of junior doctors' work patterns on the weekend: a potential contributor to the weekend effect? *Intern Med J*, 46(7): 819-825.

Douglas HE, Raban MZ, **Walter SR**, Westbrook JI. 2017. Improving our understanding of multi-tasking in healthcare: Drawing together the cognitive psychology and healthcare literature. *Appl Ergon*, 59(Part A): 45-55.

Chapter 1

Introduction

1.1 Background

1.1.1 The origins of the study of interruptions and multitasking in healthcare

Patient harm due to human error is a significant and ongoing issue in healthcare. Studies documenting the impact of error in hospitals date back over 50 years (Barr, 1955; Schimmel, 1964), but from the early 1990s large scale studies began to emerge providing startling evidence of the extent of iatrogenic harm in hospitals. One of the first such studies found that 3.7% of hospitalisations in New York State, USA, resulted in an adverse event, that is, an injury caused by medical management (Brennan et al., 1991). Of these, 27.6% were attributable to negligence, corresponding to 1.0% of all hospitalizations. A few years later a major population-based study of Australian hospitals found that 16.6% of admissions involved an adverse event with a projected annual impact of 12 000 to 23 000 deaths and a cost of 3.3 million bed days (Wilson et al., 1995). Fifty-one percent of these events were considered highly preventable, that is to say error-related, representing more than eight percent of all admissions and 1.7 million annual bed days. By way of comparison, in the year after Wilson et al.'s study there were 1 956 motor vehicle traffic accident deaths and 12 137 stroke deaths (ABS, 1993).

Since those studies there has been a considerable and growing research focus on understanding the complexities of the socio-technical systems in which healthcare is delivered and the factors that contribute to error-related harm in those settings. This focus has been bolstered more recently by the World Health Organization's development of global priorities for patient safety research (WHO, 2009; Bates et al., 2009). Much of patient safety research is about reducing negatives: minimising errors, mitigating their impacts and building defences into healthcare systems (Reason, 2000), a paradigm now referred to as Safety I (Hollnagel et al. 2015). In this vein, a seminal model of error often used in healthcare research is Reason's Swiss cheese model of system accidents (2000). This is based on the concept of a system as being composed of layers, or slices (e.g. system design, clinician training, work processes, organisational culture), each of which offers some defence against errors. In any one layer an error may occur, a hole in a slice so to speak, but harm only occurs when holes in multiple slices align.

A fundamental layer of the healthcare system is the clinical work process, and disruptions to this process have been identified as a potential element in error production in healthcare. In the course of treating patients, clinicians perform sequences of tasks and activities to juggle multiple competing demands (Patterson, Ebright et al., 2011). This can often involve interleaving tasks related to different patients (Ebright et al., 2003; Lange et al., 2016), performing multiple tasks at once and frequently having to respond to interruptions (Chisholm et al., 2000; Westbrook et al., 2008; Westbrook et al., 2011). The first of these work patterns corresponds to interleaved multitasking, sometimes called task-switching in the psychology literature, while the second is termed concurrent multitasking (Douglas et al., 2017). The third work pattern is generally referred to as interruption, although other terms such as break in task (Chisholm et al., 2000) have been used. Interruption can be considered a special case of task-switching where the switch from one task to another is prompted by some external event. These aspects of clinical practice represent the way clinicians manage work through the prioritisation, timing and sequencing of tasks. Many studies have focused on isolated aspects of the process - interruptions or multitasking alone - however, the work in this thesis is based on a more complete view of clinical work as a continuous process evolving through time, within which events such as interruption and multitasking occur.

Interruptions and both interleaved and concurrent multitasking have been studied extensively in many domains: cognitive psychology, human-computer interaction, aviation, organisational psychology, industrial engineering, and human factors and ergonomics. An online repository for research on interruptions lists over 800 publications (interruptions.net/literature.htm). In particular, research from aviation and the experimental fields of psychology and human-computer interaction (HCI) tells us that these aspects of a work process carry with them the risk of error, that is, they can potentially create holes in a system's cheese slices. It was largely studies from these domains that influenced the push to examine such phenomena in healthcare.

From experimental studies there is extensive evidence that both interruptions and multitasking have negative effects on task performance. To outline some key findings, interruptions have been associated with reduced performance on complex tasks (Speier et al., 1999; Bailey and Konstan, 2006), increased sequence errors in tasks involving a series of steps (Altmann et al., 2014), increased risk of forgetting intentions (Einstein et al., 2003), increased task completion time (Bailey et al., 2000; 2001; Cutrell et al., 2000; 2001), and increased annoyance and anxiety (Bailey and Konstan, 2006). Concurrent multitasking has been linked with decreased accuracy and increased reaction time (Rohrer and Pashler, 2003; Nijboer et al. 2013), while interleaved multitasking has also been associated with decreased accuracy (Adler and Benbunan-Fich, 2012) as well as increased task completion time (Rogers and Monsell, 1995; Rubenstein et al., 2001).

Aviation is an industry that has focused on developing safe systems over many decades and the modern commercial aviation accident rate is testament to the effectiveness of this strategy (Kohn et al., 2000). Several studies identified both interruptions and multitasking in cockpits as factors contributing to error (Dismukes et al., 1998; Latorella, 1999; Loukopoupos et al., 2000). Hence, a key part of the safety improvement process was to regulate cockpit practice through standardised procedures and checklists with the intention of eliminating the potentially negative influence of unplanned interruptions and multitasking (Loukopoulos et al., 2009).

The emergence of interruption studies from aviation and experimental psychology in the 1990s, coincided with the landmark US Institute of Medicine report, To Err is Human, in 2000 (Kohn et al., 2000). This report extrapolated evidence about interruptions into the healthcare domain by explicitly identifying them as a likely component of medical error. The assumption of negative effects associated with interruptive aspects of the clinical work process has shaped much of the subsequent research in healthcare (Hopkinson and Mowinski Jennings, 2013), spurred on by further evidence from experimental science as well as other work domains (Altmann and Trafton, 2004; Czerwinski et al., 2004). Embedded in this assumption is often a sense that interruptions and multitasking may contribute to error risk by increasing cognitive load (Baethge and Rigotti, 2013, p.44). A systematic review of studies of interruptions in healthcare found that 21 of 35 studies assumed a disruptive effect (Grundgeiger and Sanderson, 2009). Despite this there is currently no evidential consensus regarding the negative effects of the disruptive aspects of clinical work.

What is becoming clear is the challenge of studying these aspects of healthcare within a multi-layered socio-technical system (Carayon et al., 2014; Werner and Holden, 2015). This highlights the strong need to develop the research approaches applied to the study of clinical work, both conceptually and methodologically, to adequately engage with the complexity inherent in healthcare settings.

1.1.2 Studies of interruption and multitasking in healthcare

The majority of studies of clinical work are from developed countries - USA, UK, Australia and Europe - and largely focus on hospital settings. Many of the early healthcare studies provided descriptive accounts of the interruptive aspects of clinical work, often with a focus on interruptive communication (Coiera, 1996; Coiera and Tombs, 1998; Coiera et al., 2002; Chisholm et al., 2000; Chisholm et al., 2001). This type of study has persisted to the present with typical summary measures including proportions of time on particular tasks, proportions of time spent multitasking and, especially, rates and types of interruptions (Alvarez and Coiera, 2005; France et al., 2005; Friedman et al., 2005; Laxisman et al., 2007; Woloshynowych et al., 2007; Brixey et al., 2008; Westbrook et al., 2008; 2009; 2010a; 2011; Weigl, 2011, Arabadzhyiska et al, 2014; Richardson et al., 2016; Lange et al., 2016). These studies have offered a picture of the general characteristics of clinical work and their variation between sites and healthcare settings, but have not provided evidence of the function of interruption and multitasking in clinical work.

A number of studies have examined associations between aspects of clinical work - predominantly interruptions - and outcomes. These outcomes include clinician-level effects such as self-reported workload and strain, task-level effects such as failure to resume an interrupted task and the time cost of resuming such tasks, and clinical outcomes such as dispensing errors or medication administration errors. These studies form the basis for our understanding of the impacts of interruption and multitasking in healthcare, however, the complexity of those settings is not necessarily reflected in their design or analysis.

The simplest of the analyses documented particular instances where an interruption was followed by poor recall of information associated with the suspended task or non-resumption of that task (Collins et al., 2007). Similarly Drews (2007) observed whether or not interruptions preceded error in an intensive care unit (ICU), and also recorded relative frequencies of action following an interruption, including immediate resumption of a task following the interruption, delayed resumption, non-resumption, and omission of steps in a sequence upon resumption. Several studies classified interruption as having either a positive, negative or neutral outcome according to study-specific definitions of those judgments (Campbell et al., 2012; McGillis Hall et al., 2008; 2010). While all of these studies attempt to link interruptions with some outcome measure, none performed any formal statistical analyses to assess the strength or magnitude of the associations.

Some authors have attempted to assess ecological associations between either interruptions or multitasking and various outcome measures. An ecological analysis relates variables that have been aggregated. In the studies discussed in this paragraph the aggregation is usually over some unit of time such as half hour intervals, work shifts or observation sessions. Some studies have assessed bivariate relationships via correlation analysis. These found that more frequent communication distractions were associated with less frequent completion of intraoperative patient checks when those measures were aggregated by operation period (Sevdalis et al., 2014), and that multitasking was associated with self-reported strain when aggregated by observed shifts (Weigl et al., 2013). More frequently ecological analyses have taken various multivariate approaches. Weigl et al. (2014) used partial correlation analysis to identify that case-irrelevant communications in operating rooms may be beneficial for reducing mental fatigue and stress in routine cases, but may contribute to surgeons' mental focus deteriorating. Weigmann et al. (2007) also used multivariate regression with five covariates related to what they defined as 'flow disruptions', finding that teamwork related disruptions were associated with increased surgical error when aggregated per operation. Weigl et al.'s (2012) multivariate analysis found that the interruption rate per shift was associated with doctors' self-rated workload after adjusting for time of day and doctor position. Baethge and Rigotti (2013) applied hierarchical linear regression to data on nurses' work aggregated by day, and found that interruptions were negatively associated with satisfaction with one's own performance, forgetting intentions and irritation. A study of an ambulatory care pharmacy used multivariate regression to investigate an association between interruption rates and dispensing error rates adjusted for measures of individual workload and distractibility (Flynn et al., 2005). The association was only significant for data aggregated into half hour intervals, but not when aggregated by prescription sets. This highlights an issue with analysing aggregated data: significance is affected by the granularity of aggregation units. Furthermore, associations at an aggregated level do not allow inference about effects of individual interruptions.

A few studies have conducted analyses at a more granular level, that is to say, on data that has not been aggregated. Grundgeiger et al. (2010) used linear regression to model the time to resume a task following an interruption (resumption lag) among nurses in the ICU of an Australian hospital. The duration of the interruption and whether the interruption required changing location were significantly associated with resumption lag. Westbrook at al. (2010b) examined the association between interruptions and medication administration errors in general wards of two Australian hospitals. Multivariate Poisson regression was applied to data at the level of individual medication administrations with covariates related to interruptions, characteristics of nurses and patients, plus a variable to differentiate the two hospitals. Interruptions were significantly associated with errors, although no adjustment was made for the fact that both of these variables tend to be proportional to time. That is to say, the counts of both interruptions and errors would, on average, increase with observation time even if there was no association. This makes it difficult to determine whether the reported association is just a corollary of this temporal correlation or whether there is a real effect of interruptions on errors.

Despite the many factors that influence an individual's work practice (Carayon et al., 2014) only some of these studies apply multivariate analyses, usually with only a small number of covariates. Of those studies, various forms of linear regression dominate including several analyses that treated rates as continuous (Flynn et al., 2005; Weigl et al., 2012) where other forms of generalised linear models, namely Poisson or Negative binomial, would have been more appropriate. Few studies have attempted to address forms of dependence between observations, such as correlation within participants (Grundgeiger et al., 2010; Baethge and Rigotti, 2013), and none attempted to allow for temporal correlation between outcome values, i.e. autocorrelation.

1.1.3 Methodological limitations of healthcare studies

After some two decades of research, the quantitative evidence of the role of the interruptive aspects of clinical work in patient safety is mixed and our understanding is far from complete. There are several factors that have limited progress in this field: 1) a restricted research focus driven by the pervasive assumption about the negative effects of interruptive aspects of clinical work, 2) variable use of terminology and imprecise definition and operationalisation of concepts, 3) limited engagement with the complexity of healthcare contexts including minimal utilisation of the many available and applicable statistical techniques, and 4) very little development of quantitative methodology to solve questions specific to the field.

1.1.3.1 Negative assumptions about interruption and multitasking

Despite the common assumption about negative effects of interruption and multitasking in healthcare, there is a considerable amount of evidence that suggests otherwise. In the experimental domain, being interrupted has been shown to increase mental arousal, decrease boredom (Speier, 1999) and to shorten task completion time with no loss of quality (Zijlstra, 1999; Mark et al., 2008). Another study showed a non-linear relationship between productivity and the frequency of task switching, with moderate levels of switching being better for productivity than low or high levels (Adler and Benbunan-Fich, 2012). Similarly in the healthcare domain these aspects of work can have a neutral or positive effect as shown in some of the studies in section 1.1.2 above, and also discussed in several review papers (Grundgeiger and Sanderson, 2009; Rivera-Rodriguez and Karsch, 2010). Garrett and Caldwell (2006) note that an interruption's status as positive or negative relates to the relative priority of the primary and incoming tasks. For example, a nurse interrupting a doctor to clarify the details of a prescription would be seen as positive if the doctor was writing patient notes, since the avoidance of a medication administration error takes priority. In contrast the same interruption while a doctor was in the midst of patient resuscitation would be seen as negative. Anthony et al. (2010) discuss the role of interruptive communication in timely information transfer, which frames this aspect of work as integral to quality care.

The design of studies around negative effects also ignores the fact that much of the negative evidence from outside the healthcare domain involved participants performing unfamiliar experimental tasks, in contrast to healthcare settings where clinicians have at least months if not years of experience during which they may hone strategies and to some degree a sense of familiarity with the typical interruptive aspects of their work. Hence, the negative assumption that has driven much of the research on clinical work may be both simplistic and narrow and may have strongly influenced study design, as well as which variables are selected and the way they are measured (Hopkinson and Mowinski Jennings, 2013). Thus there is scope for a more holistic approach to understanding the function of factors such as interruption and multitasking in patient safety without predetermined judgements as to their effects. This suggests a need for exploratory or hypothesis generating research that can subsequently inform more targeted testing of specific hypotheses. It also prompts rethinking of study design, both in terms of conceptualisation of clinical work and the selection of factors to be observed and analysed.

1.1.3.2 Conceptualising the clinical work process

The quantitative study of clinical work relies on the way aspects of the work process are defined and the ways in which those definitions are operationalised during data collection. In the healthcare literature, multiple concepts have been defined under the same term and multiple terms have been applied to the same concept. This has resulted in pervasive inconsistency and imprecision in both terminology and definitions (Brixey et al., 2007; Biron et al., 2009; Sasangohar et al., 2012). This not only limits the generalisability of study results, but makes comparison between them difficult (Grundgeiger and Sanderson, 2009). Several authors have proposed standardisation of definitions (McFarlane, 1997; Brixey et al., 2007; Coiera, 2012) although there is still no consensus. This is potentially due to the operationalisation of definitions necessarily being specific to each new setting in which they are applied.

In controlled experiments, studies can be designed so that participant behaviour fits clear definitions of interruptions and multitasking. Conversely, in uncontrolled healthcare settings definitions must be created to fit participant behaviour. Clinicians manage multiple demands in many creative ways such that behaviours that might be considered interruption or multitasking according to experimental definitions, can occur in unlikely sequences or may be mixed together in ambiguous ways. This means that definitions used in experimental studies can be overly simple for studying work in healthcare settings. Hence there is a need for context specific definitions of the clinical work process that encompass the many scenarios that can occur. There is also a need for precise operationalisation of definitions so that observers can classify the myriad of work actions in a consistent way; something that is rarely transparently reported in healthcare studies.

1.1.3.3 The complexity of work in healthcare contexts

Much of the foundational evidence on interruption and multitasking has been generated from the isolated environments of the aeroplane cockpit and psychology computer lab. These settings are vastly simpler than the context in which healthcare takes place, particularly when compared to critical care settings such as the emergency department (ED) where workload and casemix are highly variable (Levin, 2006). Several authors have noted the need to move beyond descriptive accounts of clinical work (Grundgeiger and Sanderson, 2009; Coiera, 2012; Hopkinson and Mowinkski-Jennings, 2013), but it is only in recent years that researchers are starting to recognise the extent of the complexity of healthcare work systems (Carayon et al., 2014; Werner and Holden, 2015). Rittel and Webber's concept of a 'wicked problem' (1973) has been used to describe the study of healthcare in that it presents problems that are "dynamic with multiple sets of complex, interacting issues that evolve in an emergent social context" (Westbrook et al., 2007, p.747). While we must acknowledge the limits of quantitatively studying complex socially contingent settings, there is, however, considerable scope for quantitative design and analysis to complement studies from other research paradigms. There are now well honed direct observational techniques for capturing the individual tasks and interactions involved in clinical work (Westbrook et al., 2009; Lopetegui et al., 2014) and much of the work in this thesis will focus on data collected in this way.

A human factors model of patient safety specifies aspects of the health system at the levels of the person, tasks, physical environment and the organisation (Carayon et al., 2014). In line with this it is possible to record characteristics of clinicians, the work environment, the team, as well as time dependent factors such as workload, fatigue and stress. This can be combined with information on the tasks and interactions that make up the work process to build a more comprehensive account of the many factors influencing clinical work at multiple levels. To make sense of such data, there are many existing multivariate modelling techniques that can support the disentangling of interrelationships at different levels, and these methods have been barely explored in the study of clinical work to date.

1.1.3.4 Development of statistical methodology

The unique complexity of quantitatively studying clinical work prompts many opportunities for developing bespoke statistical methodology. To date only one paper has proposed a new method to solve a problem specific to the field: Brown and Dunsmuir (2010) devised a technique to assess whether interruptions had an impact on the time taken to complete a task, something that could not previously be assessed with data from direct observations of clinical work. The key problem addressed in that work was the fact that the number of interruptions during a task is proportional to task length. That is, even if interruptions have no effect on task completion time, there will still be a correlation between these two measures. This is referred to as length bias and is pervasive in the study of clinical work. Rarely has this effect been recognised or acknowledged, leading to results about interruptions being reported as true effects when they are potentially just reflections of length bias (Trbovich et al., 2010; Weigmann et al., 2007; Westbrook et al. 2010b). Hence there is a need for statistical techniques aimed at this issue to enable researchers to separate genuine effects from artefacts of length bias.

1.2 Thesis Aims

The overarching aim of this work was to promote improved understanding of the patient safety implications of the clinical work process through advancing and applying quantitative observational methodology. To achieve this aim my research focused on four objectives:

1. To critically review the methodology used to date in quantitative observational studies of healthcare and identify ways in which the design, data collection and analysis can be improved from current practice.

2. To develop a context-appropriate conceptualisation of the clinical work process.

3. To advance statistical approaches for the analysis of complex observational data through innovative application of existing methods and development of new techniques.

4. To generate new evidence regarding everyday clinical work practices through the application of these conceptual and statistical developments in real world hospital settings.

1.3 Thesis structure

The core of this thesis comprises five publications. Two of these are peer-reviewed journal articles, two currently under review with scientific journals and one is a peer-reviewed conference paper. Following on from points foreshadowed in this chapter, the article in chapter 2 expands the discussion of the methodological limitations facing the field of observational studies of clinical work. This publication delineates the research space within which the four subsequent papers tackle those methodological issues.

As outlined above, terminology and conceptualisation of clinical work are major limitations for the field. Chapter 3 gives a brief discussion of the schema of concepts and terms that are subsequently applied in chapters 4 to 7. In any one of these papers there is only scope to describe those aspects relevant to the individual analysis. Thus chapter 3 is included to provide a complete account of the conceptual framework.

Chapters 4 and 5 are empirical studies, the first of which uses an exploratory statistical modelling approach on pre-existing data from three previous studies. Chapter 5 represents a prospective application of the concepts outlined in chapter 3 in that the study was designed around them. Details of the study protocol are provided in the appendix. Chapter 6 describes the development of a new statistical technique to assess the impact of interruptive events on the time to complete clinical tasks. This is followed in chapter 7 by an application of this method to observational data on doctors in several hospital settings. The eighth and final chapter then provides an overarching discussion of the contribution of the thesis and the implications for both clinical practice and further research.

References

Adler RF, Benbunan-Fich R. 2012. Juggling on a high wire: Multitasking effects on performance. *Int J Hum Comput St*, 70: 156-168.

Altmann EM, Trafton JG. 2004. Task interruption: resumption lag and the role of cues. In: 26th Annual Meeting of the Cognitive Science Society. Chicago.

Altmann EM, Trafton JG, Humbrick DZ. 2014. Momentary interruptions can derail the train of thought. *J Exp Psychol Gen*, 143(1); 215-226.

Alvarez G, Coiera E. 2005. Interruptive communication patterns in the intensive care unit ward round. *Int J Med Inform*, 74: 791-796.

Arabadzhiyska PN, Baysari MT, Walter SR, Day RO, Westbrook JI. 2013. Shedding light on junior doctors' work practices after hours. *Intern Med J*, 43(12): 1321-1326.

Australian Bureau of Statistics. Causes of Death, Australia, 1993. Catalogue number 3303.0. ABS: Canberra, 1994.

Baethge A, Rigotti T. 2013. Interruptions to workflow: Their relationship with irritation and satisfaction with performance, and the mediating roles of time pressure and mental demands. *Work and Stress*, 27(1):43-63.

Bailey BP, Konstan JA. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Comput Human Behav*, 22(4): 685-708.

Bailey BP, Konstan JA, Carlis JV. 2000. Measuring the effects of interruptions on task performance in the user interface. In: IEEE Conference on Systems, Man, and Cybernetics 2000, IEEE: 757-762.

Bailey BP, Konstan JA, Carlis JV. 2001. The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In: Human-Computer Interaction INTERACT 2001 Conference Proceedings. IOS Press, IFIP: 593-601.

Barr DP. 1955. Hazards of modern diagnosis and therapy - The price we pay. *J Am Med Assoc*, 159(15): 1452-1456.

Bates DW, Larizgoitia I, Prasopa-Plaizier N, Jha AK. 2009. Global priorities for patient safety research. *Brit Med J*, 338(7704): 1242-1244.

Biron AD, Loiselle CG, Lavoie-Tremblay M. 2009. Work interruptions and their contribution to medication administration errors: an evidence review. *Worldviews Evid Based Nurs*, 6(2): 70-86.

Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, Newhouse JP, Weiler PC, Hiatt HH. 1991. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study I. *New Eng J Med*, 324: 370-376.

Brixey JJ, Robinson DJ, Johnson CW, Johnson TR, Turley JP, Zhang J. 2007. A concept analysis of the phenomenon interruption. *Adv Nurs Sci*, 30(1): E26-E43.

Brixey JJ, Tang Z, Robinson DJ, Johnson CW, Johnson TR, Turley JP, Patel VL, Zhang J. 2008. Interruptions in a level one trauma center: A case study. *Int J Med Inform*, 77: 235-241.

Brown BM, Dunsmuir WTM. 2010. Adjusting for length biased sampling when assessing the impact of interruptions on total time on task in a hospital emergency department setting. *Qual Saf Health Care*, 19: 284-289, online only supplement.

Campbell G, Arfanis K, Smith AF. 2012. Distraction and interruption in anaesthetic practice. *Brit J Anaesth*, 109(5): 707-715.

Carayon P, Wetterneck TB, Rivera-Rodriguez AJ, Schoofs Hundt A, Hoonakker P, Holden R, Gurses AP. 2014. Human factors systems approach to healthcare quality and patient safety. *Appl Ergon*, 45: 14-25.

Chisholm CD, Collison EK, Nelson DR, Cordell WH. 2000. Emergency department workplace interruptions: Are emergency physicians “interrupt-driven” and “multitasking”? *Acad Emerg Med*, 7(11): 1239-1243.

Chisholm CD, Dornfeld AM, Nelson DR, Cordell WH. 2001. Work interrupted: a comparison of workplace interruptions in emergency departments and primary care offices. *Ann Emerg Med*, 38(2): 146-151.

Coiera E. 1996. Clinical communication: a new informatics paradigm. Proc AMIA Annu Fall Symp 1996:17-21.

Coiera E. 2012. The science of interruption. *BMJ Qual Saf*, 21(5): 357-360.

Coiera E, Jayasuriya R, Hardy J, Bannan A, Thorpe MEC. 2002. Communication loads on clinicians in the emergency department. *Med J Aust*, 176: 415-418.

Coiera E, Tombs V. 1998. Communication behaviours in a hospital setting: an observational study. *Brit Med J*, 316: 673-677.

Collins SA, Currie L,Patel VL, Bakken S, Cimino JJ. 2007. Multitasking by clinicians in the context of CPOE and CIS use. In: Congress Proceedings MedInfo 2007, Brisbane, Australia: 958-962.

Cutrell EB, Czerwinski M, Horvitz E. 2000. Effects of instant messaging interruptions on computing tasks. In: Proceedings of the CHI 2000 conference on Human factors in computing systems, Extended Abstracts. New York: ACM Press: 99-100.

Cutrell EB, Czerwinski M, Horvitz E. 2001. Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In: Human-Computer Interaction INTERACT 2001 Conference Proceedings. IOS Press, IFIP: 263-269.

Czerwinski M, Horvitz E, Wilhite S. 2004. A diary study of task switching and interruptions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI 2004. ACM. Vienna, Austria, pp. 175-182.

Dismukes RK, Young G, Sumwalt R. 1998. Cockpit interruptions and distractions: effective management requires a careful balancing act. ASRS Directline.

Douglas HE, Raban MZ, Walter SR, Westbrook JI. 2017. Improving our understanding of multi-tasking in healthcare: Drawing together the cognitive psychology and healthcare literature. *Appl Ergon*, 59(Part A): 45-55.

Drews FA. 2007. The frequency and impact of task interruptions on patient safety in the ICU. In: Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting: 683-686.

Einstein GO, McDaniel MA, Williford CL, Pagan JL, Dismukes RK. 2003. Forgetting of intentions in demanding situations is rapid. *J Expl Psychol Appl*, 9(3): 147-162.

Ebright PR, Patterson ES, Chalko BA, Render ML. 2003. Understanding the complexity of registered nurse work in acute care settings. *J Nurs Admin*, 33(12): 630-638.

Flynn EA, Barker KN, Gibson JT, Pearson RE, Berger BA, Smith LA. 1999. Impact of interruptions and distractions on dispensing errors in an ambulatory care pharmacy. Am *J Health Syst Pharm*, 56 (13): 1319-1325.

France DJ, Levin S, Hemphill R, Chen K, Rickard D, Makowski R, Jones I, Aronsky D. 2005. Emergency physicians' behaviors and workload in the presence of an electronic whiteboard. *Int J Med Inform*, 74: 827-837.

Friedman SM, Elinson R, Arenovich T. 2005. A study of emergency physician work and communication: A human factors approach. *Isr J Emerg Med*, 5(3): 35-42.

Grundgeiger T, Sanderson P. 2009. Interruptions in healthcare: theoretical views. *Int J Med Inform*, 78(5): 293-307.

Grundgeiger T, Sanderson P, MacDougall HG, Venkatesh B. 2010. Interruption management in the intensive care unit: predicting resumption times and assessing distributed support. *J Exp Psych Appl*, 16(4): 317-334.

Hollnagel E, Wears RL, Braithwaite J. From Safety-I to Safety-II: A White Paper. The Resilient Health Care Net: 2015. URL: http://resilienthealthcare.net/onewebmedia/WhitePaperFinal.pdf. Accessed: 24 Nov 2015.

Hopkinson SG, Mowinski Jennings B. 2013. Interruptions during nurses' work: A state-of-the-science review. *Res Nurs Health*, 36: 38-53.

Horst WJ, Rittel HWJ, Melvin M, Webber MM. 1973. Dilemmas in a General Theory of Planning. *Policy Sciences*, 4: 155-169.

Interruptions: Literature. http://interruptions.net Accessed: 8th March 2016.

Kohn LT, Corrigan JM, Donaldson MS (Eds.). 2000. To err is human: building a Safer Health System (Vol. 6). National Academies Press: Washington DC, USA.

Lange K, Nowak M, Zoller R, Lauer W. 2016. Boundary conditions for safe detection of clinical alarms: An observational study to identify the cognitive and perceptual demands in an Intensive Care Unit. In: de Waard D, Brookhuis KA, Toffetti A, Stuiver A, Weikert C, Coelho D, Manzey D, Unal AB, Rttger S, Merat N (Eds.) 2016. Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959. Available from http://hfeseurope.org.

Latorella KA. 1999. Investigating interruptions: Implications for flightdeck performance. Hampton VA: Langley Research Center, National Aeronautics and Space Administration.

Laxisman A, Hakimzada F, Sayan OR, Green RA, Zhang J, Patel VA. 2007. The multitasking clinician: Decision-making and cognitive demand during and after team handoffs in emergency care. *Int J Med Inform*, 76: 801-811.

Levin S, France DJ, Hemphill R, Jones I, Chen KY, Rickard D, Makowski R, Aronsky D. 2006. Tracking workload in the emergency department. *Hum Factors*, 48: 526-539.

Lopetegui M, Yen PY, Lai A, Jeffries J, Embi P, Payne P. 2014. Time motion studies in healthcare: what are we talking about? *J Biomed Inform*, 49: 292-299.

Loukopoulos LD, Dismukes RK, Barshi I. 2001. Cockpit interruptions and distractions: A line observation study. In: The 11th International Symposium on Aviation Psychology. Columbus, OH, USA.

Loukopoulos, L. D., Dismukes, R. K., Barshi, I. 2009. The multitasking myth. Handling complexity in real-world operations. Ashgate Publishing Limited: Farnham, UK.

Mark G, Gudith D, Klocke U. 2008. The cost of interrupted work: more speed and stress. CHI 2008 Proceedings, April 5-10, 2008 Florence, Italy.

McFarlane DC. 1997. Interruption of people in human-computer interaction: a general unifying definition of human interruption and taxonomy. Naval Research Laboratory: Washington, DC, USA.

McGillis Hall L, Pedersen C, Hubley P, Ptack E, Hemingway A, Watson C, Keatings M. 2008. Interruptions and pediatric patient safety. *J Pediatr Nurs*; 25(3): 165-75.

McGillis Hall L, Ferguson-Paré MA, Peter E, White D, Besner J, Chisholm A, Ferris E, Fryers M, Macleod M, Mildon B, Pedersen C. 2010. Going blank: factors contributing to interruptions to nurses' work and related outcomes. *J Nurs Manag*, 18(8): 1040-1047.

Nijboer M, Taatgen NA, Brands A, Borst JP, van Rijn H. 2013. Decision making in concurrent multitasking: Do people adapt to task interference? *PLoS One*, 8(11): e79583.

Patterson ES, Ebright PR, Saleem JJ. 2011. Investigating stacking: How do registered nurses prioritize their activities in real-time? *Int J Ind Ergonom*, 41: 389-393.

Reason J. 2000. Human error: Models and management. *Brit Med J*, 320(7237): 768-770.

Richardson LC, Lehnbom EC, Baysari MT, Walter SR, Day RO, Westbrook JI. 2016. A time and motion study of junior doctors' work patterns on the weekend: a potential contributor to the weekend effect? *Intern Med J*, 46(7): 819-825.

Rivera-Rodriguez AJ, Karsh BT. 2010. Interruptions and distractions in healthcare: review and reappraisal. *Qual Saf Health Care*, 19: 304-312.

Rogers RD, Monsell S. 1995. Costs of a predictable switch between simple cognitive tasks. *J Exp Psychol Gen*, 124: 207-231.

Rohrer D, Pashler H. 2003. Concurrent task effects on memory retrieval. *Psychon B Rev*, 10: 96-103.

Rubinstein JS, Meyer DE, Evans JE. 2001. Executive control of cognitive processes in task switching. *J Exp Psychol Human*, 27(4): 763-797.

Sasangohar F, Donmez B, Trbovich P, Easty AC. 2012. Not all interruptions are created equal: positive interruptions in healthcare. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2012, 56: 824-828.

Schimmel EM. 1964. The hazards of hospitalization. *Ann Intern Med*, 60: 100-110.

Sevdalis N, Undre S, McDermott J, Giddie J, Diner L, Smith G. 2014. Impact of intraoperative distractions on patient safety: a prospective descriptive study using validated instruments. *World J Surg*, 38(4): 751-758.

Speier C, Valacich JS, Vessey I. 1999. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, 30(2): 337-360.

Trbovich P, Prakash V, Stewart J, Trip K, Savage P. 2010. Interruptions during the delivery of high-risk medications. *J Nurs Admin*, 40(5): 211-218.

Weigl M, Antoniadis S, Chiapponi C, Bruns C, Sevdalis N. 2014. The impact of intra-operative interruptions on surgeons' perceived workload: an observational study in elective general and orthopedic surgery. *Surg Endosc*, 29(1): 145-53.

Weigl M, Muller A, Sevdalis N, Angerer P. 2013. Relationships of multitasking, physicians' strain, and performance: an observational study in ward physicians. *J Patient Saf*, 9: 18-23.

Weigl M, Muller A, Vincent C, Angerer P, Sevdalis N. 2012. The association of workflow interruptions and hospital doctors' workload: a prospective observational study. *BMJ Qual Saf*, 21: 399-407.

Weigl M, Muller A, Zupanc A, Glaser J, Angerer P. 2011. Hospital doctors' workflow interruptions and activities: an observational study. *BMJ Qual Saf*, 20: 491-497.

Wiegmann DA, ElBardissi AW, Dearani JA, Daly RC, Sundt TM. 2007. Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. *Surgery*, 142(5): 658-665.

Westbrook JI, Ampt A. 2009. Design, application and testing of the Work Observation Method by Activity Timing (WOMBAT) to measure clinicians' patterns of work and communication. *Int J Med Inform*, 78: S25-S33.

Westbrook JI, Ampt A, Kearney L, Rob MI. 2008. All in a day's work: an observational study to quantify how and with whom doctors on hospital wards spend their time. *Med J Aust*, 188(9):506-509.

Westbrook JI, Braithwaite J, Georgiou A, Ampt A, Creswick N, Coiera E, Iedema R. 2007. Multimethod Evaluation of Information and Communication Technologies in Health in the Context of Wicked Problems and Sociotechnical Theory. *J Am Med Inform Assoc*, 14: 746-755.

Westbrook JI, Coiera E, Dunsmuir WTM, Brown BM, Kelk N, Paoloni R, Tran C. 2010a. The impact of interruptions on clinical task completion. *Qual Saf Health Care*, 19(4): 284-289.

Westbrook JI, Duffield C, Li L, Creswick NJ. 2011. How much time do nurses have for patients? A longitudinal study quantifying hospital nurses' patterns of task time distribution and interactions with health professionals. *BMC Health Serv Res*,11(319).

Westbrook JI, Woods A, Rob MI, Dunsmuir WTM, Day RO. 2010b. Association of interruptions with an increased risk and severity of medication administration errors. *Arch Intern Med*, 170(8): 683-690.

Wilson M, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. 1995. The quality in Australian health care study. *Med J Aust*, 163: 458-71.

Woloshynowych M, Davis R, Brown R, Vincent C. 2007. Communication Patterns in a UK Emergency Department. *Ann Emerg Med*, 50: 407-413.

World Health Organization. 2009. Global priorities for patient safety research. WHO: Geneva, Switzerland.

Zijlstra FRH, Roe RA. 1999. Temporal factors in mental work: effects of interrupted activities. *J Occup Organ Psych*, 72: 163-185.

Chapter 2

Methodological challenges for *in situ* observational studies

2.1 Chapter background

The article in this chapter reviews and discusses the methodological aspects of quantitative observational studies of work processes, with a particular focus on clinical work. There has been a number of relatively recent reviews of largely quantitative studies on interruptions in healthcare (Grundgeiger and Sanderson, 2009; Li et al., 2012; Rivera-Rodriguez and Karsh, 2010; Biron, 2009; Hopkinson and Mowinski Jennings, 2013). However, there has been limited discussion of observational and statistical methods, despite the fundamental importance of appropriate methodology for studies in such complex settings. This paper identifies key areas in need of methodological improvement, makes explicit aspects of the complexity of clinical settings, and proposes ways forward in terms of methodology to accommodate that complexity. In addition to providing a roadmap for this thesis, the intention of the paper was also to outline future methodologically directions for the general field of quantitative observational studies of clinical work.

The article was selecteded for publication in a special issue of the International Journal of Human Computer Studies on interruption and multitasking. It directly addresses objective 1 of the thesis, but also lays the foundation for the other objectives. A fuller discussion of the points outlined in the introductory chapter is given in a way that makes the concepts applicable to work processes in general, not just healthcare. Also this paper puts forward the case for the use of a particular type of observational study, the workflow time study, as a vehicle for advancing our understanding of clinical work and as an approach that can support many of the proposed methodological advances. The data analysed in subsequent chapters was collected via this observational approach.

In parallel to writing this paper, the STRATOS initiative (STRengthening Analytical Thinking for Observational Studies) was being developed. This initiative aimed to highlight the need for greater statistical rigour in observational studies generally, and aimed to produce guidance documents for best practice in the statistical method of observational studies (Sauerbrei et al., 2014). The thrust of the paper in this chapter has parallels with the STRATOS view, but with a specific focus on the direct observational study of clinical work. Many of the points raised below, and developed independently, align with the key STRATOS focus areas including multivariate modelling approaches, measurement error and misclassification, study design, and causal inference. The scope for methodological development in the study of clinical work is formidable, and subsequent chapters in this thesis aim to address the most pertinent of these.

References

Biron AD, Loiselle CG, Lavoie-Tremblay M. 2009. Work interruptions and their contribution to medication administration errors: an evidence review. *Worldviews Evid Based Nurs*, 6(2): 70-86.

Grundgeiger T, Sanderson P. 2009. Interruptions in healthcare: theoretical views. *Int J Med Inform*, 78(5): 293-307.

Hopkinson SG, Mowinski Jennings B. 2013. Interruptions during nurses' work: A state-of-the-science review. *Res Nurs Health*, 36: 38-53.

Li SYW, Magrabi F, Coiera E. 2012. A systematic review of the psychological literature on interruption and its patient safety implications. *J Am Med Inform Assoc*, 19(1): 6-12.

Rivera-Rodriguez AJ, Karsh BT. 2010. Interruptions and distractions in healthcare: review and reappraisal. *Qual Saf Health Care*, 19: 304-312.

Sauerbrei W, Abrahamowicz M, Altman DG, le Cessie S, Carpenter J. 2014. STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative. *Stat Med*, 33: 5413-5432.

2.2 Studying interruptions and multitasking in situ: The untapped potential of quantitative observational studies

Author contributions

The overall scope of the paper was developed by all authors. SRW wrote the first draft and led subsequent revisions. Both WTMD and JIW provided critical review on all versions.

Studying interruptions and multitasking in situ: The untapped potential of quantitative observational studies

Scott R. Walter[a], William T.M. Dunsmuir[b], Johanna I. Westbrook[a]

[a] *Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia*

[b] *School of Mathematics and Statistics, Faculty of Science, University of New South Wales, Sydney, Australia*

Abstract

Much of the large and growing body of literature on interruption and multitasking is motivated, in part, by a desire to reduce their negative effects in occupational settings, particularly those that are safety critical. Much of the existing knowledge has come from experimental studies, however, these do not necessarily generalize to non-experimental contexts. By virtue of being *in situ*, the results of observational studies are more generalizable, but internal validity remains an issue. Since many of the quantitative observational studies of interruption or multitasking to date have been largely descriptive, their full potential to contribute knowledge that informs practical improvements has been underutilized. We discuss ways to address threats to internal validity in quantitative observational studies through appropriate analysis with particular reference to workflow time studies, a form of direct observation. We also discuss the potential for more sophisticated analysis methods to both address some of the threats to internal validity and to provide more nuanced insights into the role and impacts of interruption and multitasking. In this way observational studies can contribute unique evidence to facilitate practical improvements to work practices and systems.

1. Introduction

A key motivation to understand interruptions and multitasking is to improve the accuracy and efficiency of work in occupational contexts. This is particularly true in safety critical settings such as air traffic control, aviation, healthcare, industrial process monitoring, and driving where error and inefficiency can have injurious or costly repercussions. In-depth knowledge of the role and impacts of interruptions and multitasking can inform improvements to workplace safety, practices and systems. Due to the complexity and heterogeneity of workflow and individuals in such settings, studying aspects of human work processes, such as interruptions or multitasking, presents many challenges for quantitative study design and analysis.

Several approaches can be employed to study work processes including controlled experiments, computer simulation studies, and observational studies. Both experiments and simulations can be designed to control known and unknown sources of bias and thus achieve a high level of internal validity. However, the generalizability of results is limited by their similarity to non-experimental occupational settings, that is, they can lack sufficient external validity (Shadish et al., 2002). Some experimental studies have attempted to replicate interruptions or multitasking in contexts of interest, such as an office environment (Mark et al., 2008), cockpit (Latorella, 1999), motor vehicle (Watson and Strayer, 2010) or operating room (Liu et al., 2009); however, this becomes increasingly difficult for more complex and unpredictable settings such as hospital emergency departments (ED). Computer simulation studies provide a means to model interruptions or multitasking in more complex scenarios in a controlled way [see for example: (Lebiere et al., 2001; Sierhuis et al., 2007)], but this approach is limited by the accuracy of the necessary assumptions and, as with experiments, it can also be difficult to capture all the complexities of an uncontrolled setting. To date simulation studies of work in complex settings like EDs have focused on aspects such as patient flow and staffing, but not on interruptions or multitasking - an exception being a study (Gunal and Pidd, 2006) that simulated the effect of multitasking, in the sense of concurrent patient management, on departmental performance.

There are many types of observational studies that can be applied to investigate interruption and multitasking. Qualitative observational studies can provide insights about relationships, social dynamics and individual motivations and thought processes in a way that quantitative studies cannot, and this can be valuable when studying complex socio-technical settings. Nugus and Braithwaite (2010) used an ethnographic approach in an ED to understand the seemingly opposing factors of quality and organizational efficiency: a question which encompasses issues around multitasking and interruptions. Colligan and Bass (2012) used a combination of semi-structured interviews and direct observation to examine strategies that nurses used to handle interruptions.

While all types of study can contribute important knowledge about interruption and multitasking, in this article we focus on quantitative observational studies for several reasons. They can be conducted in the setting of interest, hence making their results generalizable to at least that context or others that are similar (Black, 1996). For example, a study of medication administration errors found that the risk and severity of error increased with the number of times the administration was interrupted (Westbrook et al., 2010b). Observing interruptions of nurses *in situ* provides a more accurate assessment of their potential impact on nurses' work than results from experiments or simulations. There may also be ethical constraints on conducting experiments or interventions in safety critical settings where the effect of unintended negative consequences could be serious. The same restriction is less of an issue for observational studies where the data collection process aims to have minimal impact on the context under study. However, a major drawback to the quantitative observational approach is that it can be difficult to establish internal validity and to date this has proven restrictive to the rate of knowledge generation about interruption and multitasking, particularly in healthcare.

The majority of quantitative observational studies of interruptions or multitasking are situated in medical contexts and, as noted previously (Coiera, 2012; Grundgeiger and Sanderson, 2009), most of these have essentially taken a counting approach by simply summarising counts, rates and proportions. A select few healthcare studies have taken a more advanced approach. The previously mentioned medication administration study used a multivariate analysis to find an association between interruption and error (Westbrook et al., 2010b), while another study of intensive care unit staff used eye tracker technology and a multilevel multivariate model to analyze resumption lag following an interruption (Grundgeiger et al., 2010).

While the quantitative observational approach is well suited to healthcare, it is also applicable in other domains. Several studies of information workers have used this approach to examine concurrent task management (Czerwinski et al., 2004; Gonzalez and Mark, 2004), and Loukopoulos (2001) conducted a study of interruption and task interleaving among pilots by observing their activities from the cockpit jumpseat. In an observational study of drivers, Strayer and Drews (2006) assessed the association between concurrent hand held cell phone use while driving and failure to stop at an intersection.

The need to advance the research agenda for interruptions and multitasking in healthcare has been recently noted (Westbrook, 2014), and there is clearly considerable scope for more rigorous observational studies to contribute practically useful knowledge to occupational domains, whether healthcare or otherwise. In this paper we aim to expound the ways in which the design, data collection and analysis of quantitative observational studies of interruption and multitasking can be improved from current practice. In particular, we discuss fundamental issues with the internal and external validity of observational research in reference to interruption and multitasking, and the ways in which these issues can be mitigated through the application of existing statistical techniques. We also point out areas in which new statistical developments are needed and outline ways forward for each. Where possible, we illustrate these points via a hypothetical case study.

2. Workflow time studies

There are many approaches that can be employed to record an individual's work process, as discussed at length by Lopetegui et al. (2014). The workflow time study approach (Lopetegui et al., 2014) is a type of time and motion study that offers many advantages over other non-experimental methods applicable to work processes. It involves an external observer shadowing a participant and recording time-stamped information about their tasks and interactions to create a continuous record of the work process. It has its roots in Mintzberg's structured observation method (1970) and is also similar to systematic direct observation used in timed-event sequential analysis in psychology (Bakeman and Gottman, 1997; Chorney et al., 2010) in that it involves recording behaviour in an uncontrolled setting according to predefined operational definitions. The additional emphasis in workflow time studies is on capturing a continuous record of behaviour. It is distinct from an ethnographic approach where observed interaction or behaviour is categorised during the analysis phase (Atkinson and Hammersley, 1994). Workflow time studies have been applied to interruption and multitasking in the domains of healthcare (Weigl et al., 2011; Westbrook et al., 2010a), aviation (Loukopoulos et al., 2001) and human-computer interaction (Gonzalez and Mark, 2004; Mark et al., 2012; Su and Mark, 2008).

The continuous recording of data increases the potential to capture work complexity compared to work sampling or self-report approaches such as diary studies

(Mintzberg, 1970). It is also less prone to bias than work sampling (Finkler et al., 1993) or self-report. While audio or video recording can provide an accurate continuous record of a work process, these can easily capture non-participants and the need to seek consent from all those recorded can be prohibitive. In addition, workflow time studies open up the analysis possibilities to a wide range of existing techniques, each of which has the potential to provide innovative insights. Hence we focus on this observational approach and the ways in which it can minimise threats to internal validity and can broaden the scope for statistical analyses applicable to observational data on interruptions and multitasking.

3. Internal validity

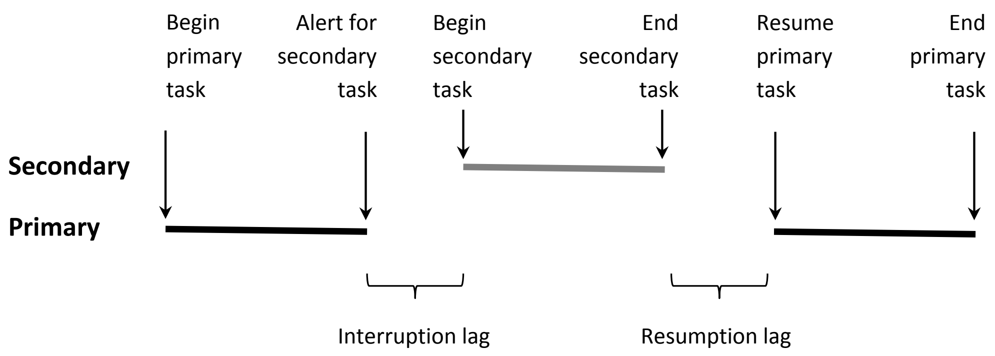
One of the main challenges in quantitative observational studies is to generate internally valid results, that is, results that are not biased. This is particularly so in complex settings where there is a network of intertwined factors at play and separating out the influences of particular factors requires addressing the many threats to internal validity. In this section we outline some of those threats and how they can be mitigated with reference to workflow time studies.

3.1 Defining interruptions and multitasking

There is much heterogeneity in the definitions of interruptions and multitasking. Many studies provide no explicit definition, while others attempt to bring some precision to particular terms, such as Trafton et al.'s (2003) often cited 'anatomy of an interruption' (Figure 1). The study of interruption and multitasking is now beset with inconsistency, with some terms having been defined to have several different meanings, and some concepts described by several different terms. For example, with reference to Trafton et al.'s model, an *interruption* has been defined as the alert for the secondary task (Chisholm et al., 2001; Czerwinski et al., 2004; Mache et al., 2012), the secondary task itself (Li et al., 2012), or the whole sequence depicted in Figure 1 (Boehm-Davis and Remington, 2009; Weigl et al., 2011). *Multitasking* has also been defined in several different ways. The notions of *concurrent* multitasking (also called *dual task performance*), *interleaved* multitasking (also called *task-switching*) and *sequential* multitasking (Adler and Benbunan-Fich, 2012; Loukopoulos et al., 2009) have been combined under a unifying theory in which these concepts represent different places on a continuum depending on the rate of switching between tasks (Salvucci et al., 2009). Yet to complicate things, task-switching that is externally triggered is sometimes called multitasking in the experimental literature [e.g. (Katidioti and Taatgen, 2014)] while generally called interruption in the healthcare literature. While several papers have discussed the panoply of terms and definitions (Brixey et al., 2007; McFarlane, 1997; Sasangohar et al., 2012), there is often an assumption that a single definition is both possible and desirable [see for example (McFarlane, 1997)].

As a way through the semantic imbroglio there are several considerations for future observational studies. First, definitions should be developed specific to the context and the research hypotheses. In complex non-experimental settings people juggle competing demands in a wide range of creative ways such that behaviours that might be considered 'interruption', 'task-switching' or 'multitasking' according to previous definitions, can occur in unlikely sequences or may be mixed together in ambiguous ways. While there are many proponents of a universal definition of interruption (Brixey et al., 2007; Grundgeiger and Sanderson, 2009; Sasangohar et al., 2012) or multitasking (Salvucci et al., 2009), if such definitions were possible, every time the definition was applied in a new context it would necessarily have to be reinterpreted and re-operationalized, hence defeating the purpose of a universal definition. Secondly, definitions should be as precise as possible to minimize measurement error or observer bias. In a fast moving fluid environment it is essential to have operational definitions that distinguish what is and isn't considered an interruption or a multitask as clearly as possible. This is necessary so that observers can translate observed behaviour into a record of the work process in a repeatable way (Hintze et al., 2002). This also supports transparency and comparability when publishing results. Developing and operationalizing definitions can be an iterative process by which definitions are tested (through piloting), adjusted and retested until they can be applied in a way that minimizes bias and error. A final and optional consideration is that definitions can be chosen according to an underlying construct of interest (Grundgeiger and Sanderson, 2009). We now introduce a hypothetical case study to illustrate these points, and this will be used throughout the paper to elucidate subsequent points. For the purpose of illustration, the study is somewhat simplified. A study aims to determine factors associated with non-resumption of interrupted tasks among doctors in an emergency department (ED). Using a similar idea to multiple resource theory (Wickens, 2002) the investigators hypothesize that non-resumption is a failure of prospective memory and occurs when the context is demanding enough that insufficient cognitive resources remain to recall the intention to resume a task. ED doctors often work in open departments as part of a team and there are many events that could potentially be included in the definition of an interruption. For a task to be at risk of not being resumed, it first has to be suspended prior to completion. This task-switching may be externally or internally triggered, but the researchers decide that knowing about external triggers is more informative for improving practice. Hence the study team decides to define an interruption as a switch from one task to another prior to completion of the original task, and where the switch is triggered by an external event. An external event is defined to include anything specifically directed towards the doctor including phone calls, questions, computer alerts and pager calls, but excluding equipment alarms, nearby conversations, other people's phones ringing and so on.

Having developed study-specific definitions and operationalised them, the next consideration is to assess the extent to which definitions are reliably applied.



Recreated from: Trafton, J.G., Altmann, E.M., Brock, D.P., Mintz, F.E., 2003. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. International Journal of Human Computer Studies 58, 583-603.

Figure 1. Trafton et al.'s anatomy of an interruption.

3.2 Intra- and inter-rater reliability

When data collection relies on an observer interpreting what they see and hear, there is potential for variation in how the definitions are applied from one observation period to the next for the same observer (*intra*-rater reliability), and from one observer to the next (*inter*-rater or *inter*-observer reliability). In addition to the definitional precision discussed above, reliability is usually optimized through observer training and a quantitative measure of agreement. For inter-rater reliability this simply means having two (or more) observers record data while observing the same scenarios. Intra-rater reliability is less easy to test as it ideally involves an observer of the same scenarios at different points in time.

Establishing a sufficient level of reliability requires a means to quantify it. The original inter-rater methods that arose from the field of psychology assess agreement between two raters classifying the same entity into a set number of categories [e.g. (Cohen, 1960)]. Although some of the univariate measures, such as Cohen's kappa, are often applied in observational studies of work processes (Lopetegui et al., 2013), these methods have two main limitations. The first is that they do not take into account the temporal ordering of tasks: the time stamp of each task and its place in the sequence of tasks are important considerations for reliability. Secondly they cannot assess agreement simultaneously between multiple variables. For example, a task may have several characteristics such as start time, duration, type (e.g. documentation) and interruption status and ideal agreement requires all characteristics to match. Determining whether data from two observers agrees on all of these attributes together is beyond the scope of existing methods. High kappa values do not necessarily mean there is good intra- or inter-rater reliability for all aspects of the collected data. The TimeCaT software includes a multidimensional measure of both inter- and intra-rater reliability by separately comparing total task time, total task count, click accuracy and sequence similarity (TimeCaT 3.9, 2013). Many methods exist for determining similarity between two strings of data or, equivalently, two multivariate records, including probabilistic record linkage (Herzog et al., 2007). Such methods have the potential to be adapted to quantify agreement in a way that takes into account the number of variables being simultaneously compared - perfect agreement becomes increasingly difficult when more variables have to match. In a similar way to Cohen's kappa, these approaches can also indicate the extent to which the level of agreement exceeds that expected due to chance alone. For example, probabilistic linkage could be used to identify the best matching unique record pairs and the sum of total match weights for all the 'best' pair matches aggregated into an overall score. To compare the likelihood of that score occurring by chance, a p-value for this total could then be obtained via a Monte Carlo permutation approach, that is, by recalculating the score for random shuffles of the data to generate a sampling distribution to which the original score can be compared.

These techniques can facilitate minimization of bias or error during direct observations which can then enable more accurate analysis results.

3.3 The importance of capturing covariates in uncontrolled settings

In an uncontrolled setting there may be many factors that simultaneously yet differentially influence a particular outcome of interest. In an experiment these factors are controlled to isolate a particular effect. Otherwise, where these factors are quantifiable, they can be analyzed concurrently to separate out the effects of each. There are many well established techniques for doing this, multivariate regression being one of the most well known. In this context these concurrent factors are often referred to as covariates. The workflow time study approach enables the simultaneous collection of data on many covariates. In this section we consider the importance of covariates in terms of two broad analysis approaches.

The first is the hypothesis driven approach which aims to test the effect of one or more predetermined factors on an outcome. In our case study example, as described in section 3.1, this might be the effect of an interrupting task involving patient resuscitation (a binary variable) on the risk of non-resumption of the original task (also binary). This is simplified to illustrate that a demanding interrupting task may be hypothesised to be more likely to cause non-resumption of the original task. Due to the non-randomized nature of the data, the relationship between these variables may be confounded by other factors such as the experience level of the doctor. A confounder is a variable that is separately related to exposure and outcome, and failure to account for it can result in bias in the effect of interest (Greenland et al., 2008). Of the many ways to deal with confounding, the most applicable to quantitative observational data is multivariate modelling where the effect of interest and all potential confounders are included as covariates. Methods for selecting covariates for regression adjustment are well covered elsewhere (Greenland, 1989; Schisterman et al., 2009). It is possible to try to collect information on as many confounders as possible, then apply variable selection techniques to find the most relevant factors. Alternatively, confounders may be hypothesised *a priori* and only information on those predetermined factors collected and adjusted for. Weigl et al. (2012) provide an example of this kind of analysis where they examined the association between interruptions and perceived workload among hospital doctors, while adjusting for time of day and doctor seniority as confounders.

Alternatively, an exploratory approach aims to identify factors associated with the outcome of interest as opposed to testing a particular hypothesis. The significant variables are distilled down from the set of all available variables via a model building process that aims to find the model that best explains the data [see for example, Hosmer and Lemeshow's purposeful selection of covariates, a model building process that aims to improve on automated stepwise methods (Hosmer and Lemeshow, 1999), section 5.2]. This type of analysis is relevant to complex socio-technical settings where there are many potential factors and little may be known about their interrelationships. In the case study this would apply to the question: what factors are associated with non-resumption of an interrupted task? In addition to the variables mentioned for the hypothesis driven approach, the researchers also consider characteristics of the interrupting task (type, interrupting person, duration, arrival time during primary task). Other covariates may further be constructed from the data such as the interruption rate in a time window preceding the interrupted task, or the number of non-resumed tasks accumulated to a given point in time. Grundgeiger et al. (2010) and Walter et al. (2014) each present examples of a model building approach.

3.4 Maximizing internal validity through analysis

Regression modelling, as mentioned in section 3.3, is a flexible way of analysing interruptions or multitasking situated in work processes. The possibilities are manifold, but there are certain aspects of regression that are important for minimising bias and have received limited attention in observational studies of interruption and multitasking to date. The importance of covariates adjustment has already been discussed, so we now also consider autocorrelation, clustering and unmeasured confounding. In data where the outcome is temporally ordered there is often correlation between the value of an outcome variable at a given time and previous values of that variable, known as autocorrelation. This is commonly dealt with by including autoregressive error terms or by including lagged values of the outcome as covariates in the regression model. Failure to account for autocorrelation can have serious impacts on accuracy and precision (Pollitt et al., 2012). As a simple example, in a study of factors associated with a clinician's choice to switch tasks or concurrently multitask when triggered by an external event, Walter et al. (2014) included the choice at the previous trigger as a potential covariate.

Another form of correlation is due to clustering. This occurs when outcome values show correlation within certain subgroups or clusters and represents another potential source of bias if ignored, particularly for standard error estimation (Diggle et al., 2002). There are a number of ways to account for this in a regression context with generalized estimating equations and random effect models being two common approaches. In general, a particular work unit or group will have its own practices and team culture, and individuals within each setting will have their own ways of working. Hence individuals and groups are two potential levels of clustering that may need to be addressed in a task-level analysis, although the levels of clustering will be specific to the study design and setting. Grundgeiger et al. (2010) and

Walter et al. (2014) present examples where random intercepts models were used to address clustering within individuals.

Multivariate modelling is an effective way to adjust for the effects of known confounders that involves optimising precision and confounder-related bias while avoiding over adjustment and unnecessary adjustment. Techniques also exist for taking into account the effect of unmeasured confounders (Hougaard, 1995; Lin et al., 1998). While less commonly performed, these methods provide a way to increase evidence for (or against) a causal relationship. If an estimated effect is relatively resilient to a range of assumptions about unmeasured confounding then this strengthens the evidence of a real relationship not due to other factors (Lin et al., 1998).

4. External validity

External validity is the extent to which study results are relevant to settings other than the original study setting. This is important in that the point of most studies is to generate knowledge that is generalizable. We outline two considerations in this vein: the influence of external observers on the participants and ensuring there is sufficient statistical power to detect genuine effects of interest.

4.1 Reactivity

The presence of an observer has the potential to influence the way an observed person behaves. For instance, being observed may promote productivity or better adherence to official procedures. This is often referred to as the Hawthorne effect and the presence of such an effect can introduce bias. The existence of this phenomenon in the original study of the Hawthorne Works in Chicago has been subsequently questioned or contradicted, the main issue being the lack of adjustment for other factors influencing workers' productivity. Jones (1992) performed a multivariate reanalysis of the original data from the relay assembly test room and reported no evidence of a Hawthorne effect after adjusting for other possible confounding factors. A reanalysis of the illumination experiments found no evidence of an immediate response to changes in light conditions (Levitt and List, 2011). However, there is still potential for observers to have an influence in other settings. While the observer ideally aims to watch a participant from a fly-on-the-wall perspective, it is unavoidable that they themselves become a part of the setting. Simple tactics to mitigate the potential for observer influence are to avoid interaction between observer and subject and for the observer to remain far enough away so as not to encroach on the subjects' performance of their work (Weigl et al., 2011). Also, assurance that performance is not being assessed and that any errors will be recorded in a non-identifiable way will help to minimise any sense of being under scrutiny. A period of acclimatisation, prior to beginning observations proper, can allow subjects to become somewhat comfortable with being directly observed. Conducting observation sessions throughout an extended period, several months say, may also reduce to possibility of sustained behaviour changes (Westbrook and Ampt, 2009).

4.2 Sufficient power

A precursor to having generalizable results is to have sufficient power to detect real effects or estimate the prevalence of events - such as interruptions - with sufficient precision. This is traditionally achieved through sample size calculation prior to commencement of observations. Most standard sample size formulae originate from the domain of experiments. Less straight forward methods exist that are applicable to non-experimental studies; however, in general the more factors being collected and analysed, the more challenging the sample size estimation becomes. This can be compounded in the absence of any prior knowledge or evidence to form the basis of the calculation assumptions.

In observational studies of work processes, sample size calculation is rarely mentioned, yet for the resource intensive workflow time study method this could be worthwhile. Applying a more-is-better approach may not necessarily result in increased power since an expanded sample may capture a more diverse group of subjects, or if the observations are carried out over an extended period of time temporal variation may be introduced. Capturing and accounting for these additional sources of variability can increase generalizability, but can also considerably augment the sample size required to maintain a given level of precision due to the inverse relationship between the number of parameters estimated by a regression model and the precision of the estimates. Also the context specific nature of observational studies means that there is a limit to how widely their results can be generalized.

Sample size determination methods exist for multivariate models, multi-level models and time series; however, for some types of analyses no directly applicable methods have been developed. In place of developing new methods it is possible to adapt existing methods by making some simplifying assumptions, or to apply several methods and use the most conservative estimate. Even if it is not possible to generate a precise sample size estimate it can still be worthwhile ensuring there is sufficient power to assess the main hypotheses of interest. Previous studies can be informative in determining sample size, but in the absence of prior relevant information another possibility is to carry out a period of observations or collect some pilot data, perform interim analyses and then use those results to update the sample size calculation based on the pilot effect sizes.

We illustrate how a researcher might go about determining sample size for a workflow time study by revisiting the case study. The researchers are interested in the relationship between a task not being resumed and the interruption of the task being caused by a resuscitation call, both binary variables. The analysis plan is to use logistic regression on all interrupted tasks with non-resumption as the outcome and resuscitation status of the interruption as the main covariate of interest. It is expected that the type of task and the time since task beginning to interruption will have confounding effects and will be included as additional covariates. From previous data the non-resumption rate has been estimated at 5% of interrupted tasks, interruptions affect 20% of tasks, and doctors complete 12 tasks per hour on average. The researcher starts with a simple approach that ignores the confounders and generates a sample size estimate of 1 647 tasks, which can be translated to about 168 hours. There is no prior information on the distribution of possible covariates or on the direction or size of their effects, so the researchers recalculate including the two confounders, but with a range of plausible distributions and effect sizes for each. This gives estimates from 103 to 225 hours. At most the study has resources for 200 hours of observation so it is decided to refine the multivariate calculation after obtaining estimates of confounder covariate effects from the initial 20 hours of observation. The refined calculation gives an estimate of 145 hours. Since this is lower than the initial estimate of 168, the researchers decide to take a conservative approach to ensure sufficient power, but to conserve at least some resources by collecting 170 hours of observations.

5. Scope for analytic innovation in quantitative observational studies

Workflow time studies enable a wide range of analysis methods relevant to observational studies of interruption and multitasking but such methods have been under used or not used thus far. We describe the application of a selection of these techniques in this section. These are well described elsewhere and we only provide a brief outline of each.

5.1 Linking interruption and multitasking to outcomes: Association and causation

The aim of much of the research on work processes is to examine associations between phenomena such as interruptions or multitasking and particular outcomes such as error or inefficiency. Due to the dynamic and complex nature of the settings in which interruptions and multitasking are endemic, establishing a link between a particular interruption, say, and a particular error is analytically challenging. While ideally we would like to know the precise cognitive steps that led to each error, i.e. the causal explanation, for the most part a quantitative observational approach identifies only observable factors that are potentially causative of errors, i.e. the causal description (Shadish et al., 2002).

There is much literature on what constitutes evidence for causality, including a number of suggested minimum conditions [e.g. Hill (1965)]. Three widely used conditions originating with John Stuart Mill (Cook and Campbell, 1979) are that the cause should precede the effect, the purported cause and effect are related, and other possible explanations for the relationship can be eliminated. Although causality cannot be established with any certainty, and there is no neat road map for doing so, observational studies that attempt to establish these three conditions may identify at least some of the component causes of error (Rothman et al., 2008). The point of statistical regression models is to quantify associations, hence effects identified through modelling satisfy the second condition, while adjustment for confounders helps to rule out other explanations for an observed association (third condition). Addressing the first condition requires incorporation of some measure of temporal ordering into the analysis.

A relatively simple approach to linking interruption and error, used in several studies to date, is to examine the association between rates aggregated over some time period. For example, Flynn et al. (1999) found an association between the rate of interruptions per half hour and the rate of medication dispensing errors aggregated over the same unit of time. This could similarly be applied to some aggregated measure of multitasking. While appealingly simple, we do not know whether the interruptions had anything to do with the errors, only that their occurrence rates were somewhat correlated. Many tasks may have been carried out during each half hour period, yet we have no sense that interruptions were temporally close to and preceding errors or whether there was another factor driving both. For example increased workload may amplify the frequency of interruptions and perceived excessive workload may increase the propensity for error. In terms of system design there is not enough information to indicate how to make changes to reduce errors.

Westbrook et al. (2010b) used a more targeted approach where the association between interruptions and errors was examined only during one particular type of task: medication administration. This is similar to a single duration measurement approach (Lopetegui et al., 2014). While this more clearly establishes the temporal proximity (although not ordering) of interruptions and errors, it ignores much of the work process that could also have contributed to the occurrence of both.

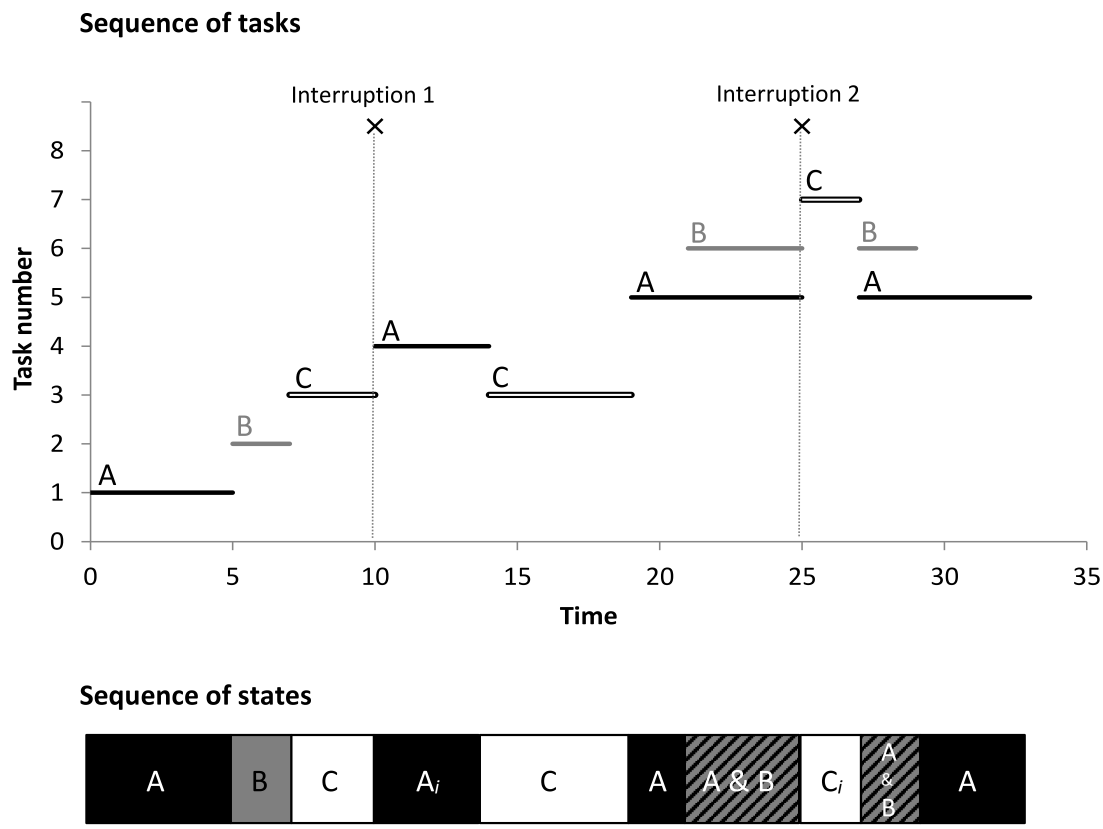
With the benefit of having a continuous record of the work process it is possible to explicitly examine temporal ordering of interruption (or multitasking) and error. This may be done by including covariates that capture information about interruptions or multitasking that precede each error. An example of this might be the interruption rate in a local time window preceding each error, or the time since last interruption. Alternatively, past or future values of a covariate may be included to examine associations at different lead or lag intervals. The effect of an interruption on reaction time has been shown to persist for a period after the interruption (Altmann and Trafton, 2007) and this is an obvious scenario in which lagged relationships could be modelled. Schildcrout and Heagarty (2005) discuss this for binary outcomes, and there is much literature on the related approach of distributed lag modelling (Almon, 1965). Since these models can also establish strength of association and adjust for other possible explanatory variables, they are a means by which a causal description can be established.

5.2 Bridging the gap between experimental psychology and observational studies

Observational studies in uncontrolled settings cannot isolate cognitive phenomena related to interruption or multitasking in the same way a controlled experiment can. One of the few studies to attempt to do so used eye tracking technology to directly measure resumption lag following interruptions (Grundgeiger et al., 2010), something only previously measured in experiments. However, in this section we broadly outline a set of modelling techniques by which inference about underlying psychological constructs (latent variables) may be drawn from directly observed behaviour. To illustrate via the case study, abandoning an interrupted task may be a failure of prospective memory and the researchers hypothesize that this is due to excess cognitive resources being consumed by the interrupting task. Hence whether the interrupted task is abandoned is considered a binary realization of the unobserved, or latent, level of available cognitive resources at that time. Similarly, a participant's choice between interruption and multitasking when triggered is hypothesized to represent the same construct, with a choice of interruption indicating less available resources than if multitasking was chosen. Other observable aspects of the work process can also be defined as realisations of the same construct to create a range of observed measures of that construct. The investigators are interested in how workload relates to the level of cognitive resources and so collect several workload measures: individual heart rate variability, department level patient load, and number of patients concurrently managed by each participant. Within the broad framework of structural equation modelling it is possible to assess the relationship between available cognitive resources and workload by modelling the observed realisations of these latent factors. As a further example of how observed behaviour can be used to make inference about latent constructs, a study of primary school children used observations of 20 different behaviours related to memory deficits and 12 validated tests of working memory to draw conclusions about the relationship between two underlying constructs: cognitive working memory and behavioural working memory (Alloway et al., 2009). The estimated correlation between the two factors was reported as 0.52. This broad approach applied to observational data of work processes may form an important means by which to study relationships among psychological phenomena related to interruption and multitasking in non-experimental settings.

5.3 Other analysis possibilities

The continuous record of a work process can be conceptualized as a series of states. For example, conversing may be considered one particular state, documenting could be another. Figure 2 provides an example illustration of how a sequence of tasks may be conceptualised as a sequence of states. This opens the way for the use of Markov models. A relatively simple possibility is then to model the probability of transitioning to various states given the current state and, optionally, given a certain number of previous states. Definitions of states can then be defined to be relevant to interruption and multitasking. For example, the probabilities of transitioning from conversing to some other task could be compared where the conversation is interrupted versus when it is not, thus capturing the effect of interruptions on workflow. There are many Markov models that may be relevant to particular hypotheses, and many relevant applications may be found in the related field of sequential behaviour analysis (Bakeman and Gottman, 1997). In one of the few examples applied to interruption and multitasking, Su and Mark (2008) use Markov transition probabilities to examine task switching (i.e. interleaved multitasking) in sequences of tasks grouped into communication chains.



Note: The sequence of tasks shows seven tasks categorised according to three task types: *A*, *B* and *C*. Two of the tasks (4 and 7) are interrupting tasks. The sequence of states uses combinations of task type and interrupting or multitasking status to define distinct states. A type *A* task that interrupts (A_i) is distinguished from a type *A* task that is not interrupting. A multitask between a type *A* and type *B* task is distinguished from a type *A* task alone or a type *B* task alone.

Figure 2. A sequence of three different task types (A, B or C) involving interruption and multitasking with an example of how the process can be conceptualised as a sequence of states.

A final analysis consideration related to the impact of particular events on task length is the phenomenon of length bias. If events such as interruptions occur at random points in time, the likelihood of a task being interrupted is proportional to its duration. Thus it is not valid to compare lengths of interrupted and uninterrupted tasks to assess whether interruptions have an effect on task length. Instead, it is necessary to estimate the length bias adjusted expected task length for a given number of interruptions, assuming there is no interruption effect, and compare this to the observed lengths. A significant difference between observed and expected values provides evidence for an interruption effect. One type of interruption effect is to lengthen the time taken to complete a task through resumption lag. This has been studied through direct timing of resumption lag in experimental studies (Altmann and Trafton, 2004) and has been assessed in an occupational setting (Grundgeiger et al., 2010); however, analytic methods to assess an interruption effect can also incorporate other types of effects such as task shortening. A method to do this was proposed by Brown and Dunsmuir (2010) as part of Westbrook et al.'s (2010a) study of emergency department doctors, although there is considerable scope for extension.

6. Discussion

We have outlined the considerable scope for quantitative observational studies to contribute important evidence about interruption and multitasking specific to occupational settings through the use of workflow time studies and increased analytic rigour and innovation. Arguably, the main motivation for studying interruption and multitasking in such settings is to improve performance by reducing error and inefficiency. The role of quantitative observational studies is manifold in this respect. They can be used to gain an understanding of the way interruption and multitasking function in a particular setting and to then inform the nature of improvements to practice. Many of the observational healthcare studies of interruption and multitasking fall into this category and such studies are necessary when little is known about a complex setting. This is analogous to the exploratory approach described in section 3.3. Observational studies can also be used to test particular hypotheses generated by previous research and related to a proposed set of changes. Further, observational studies can be applied to assess implemented changes or interventions *in situ*, that is, quasi-experimental observational studies [e.g. (Weigl et al., 2014)]. The latter two applications are more akin to the hypothesis driven approach previously described. While the risk of unintended negative effects of a poorly informed wide-spread intervention is potentially disastrous in safety critical settings, elsewhere it may be more possible to intervene in work practice at a small scale and see what happens. For example Mark et al. (2012) cut off email to a group of scientific researchers and assessed the impact on their use of interleaved multitasking. Regardless of the way observational studies are applied, the points outlined in this paper apply equally.

Where experiments can isolate specific aspects of interruption and multitasking, observational studies situated in occupational contexts can take a broader perspective. Interruptions may be a contributing factor to some negative outcomes, yet in terms of system design it may be more useful to identify what drives interruptions in the first place. If interruptions are symptomatic of high workload then a focus on managing workload may be more beneficial than focusing on interruptions alone. Improvement to a complex system may also require broadening from observations of individuals to observing the whole system, as discussed by Harr and Kaptelinin (2007). An environment characterized by interruptive communication may seem suboptimal for the individual, but could be the most efficient means of timely information transfer to ensure successful operation of the team. Conversely, reducing the level of interruptions at an individual level may not result in system wide improvement.

We have outlined many analysis techniques including models that provide insight about unobserved variables and causal relationships. However, there is a limit to what can be learned from quantitative studies of the type we have discussed. Observers can only capture a certain amount of quantitative information; hence there is a role for qualitative studies in capturing more nuanced details of interactions, as exemplified in the qualitative studies described in the introduction (Colligan and Bass, 2012; Nugus and Braithwaite, 2010). A further limitation of workflow time studies is that they can be resource intensive, with previous such studies often requiring several hundred hours of observation to capture sufficient errors and covariate information (Westbrook et al., 2010a; Westbrook et al., 2010b). This needs to be weighed against the proposed benefit of well-informed changes and the cost of poorly informed changes.

The real strength of well conducted observational research is that it can be situated in working contexts. This has the potential to provide knowledge of genuine use for improving practice, particularly in settings where the negative effects of interruption and multitasking could be costly. The many possibilities outlined in this paper underscore the untapped potential of this type of research in the study of interruption and multitasking as well as work processes in general.

Acknowledgements

This work was funded by ARC Discovery grant DP110100090 and NHMRC Program grant 1054146. SRW is funded by a NHMRC Postgraduate Scholarship. The authors would like to thank the editor, the anonymous reviewers, Dr Heather Douglas and Dr Arathi Sriprakash for their valuable feedback.

References

Adler RF, Benbunan-Fich R. 2012. Juggling on a high wire: Multitasking effects on performance. *Int J Hum Comput St*, 70(2): 156-168.

Alloway TP, Gathercole SE, Kirkwood H, Elliott J. 2009. The working memory rating scale: A classroom-based behavioral assessment of working memory. *Learn Indiv Diff*, 19(2): 242-245.

Almon S. 1965. The distributed lag between capital appropriations and expenditures. *Econometrica* 33(1),: 178-196.

Altmann EM, Trafton JG. 2004. Task interruption: Resumption lag and the role of cues. In: 26th Annual Meeting of the Cognitive Science Society, Chicago.

Altmann EM, Trafton JG. 2007. Timecourse of recovery from task interruption: Data and a model. *Psychon Bull Rev*, 14(6): 1079-1084.

Atkinson P, Hammersley M. 1994. Ethnography and participant observation. In: Denzin NK, Lincoln YS (Eds.), Handbook of qualitative research. Sage Publications, Inc., Thousand Oaks, CA, USA, pp. 248-261.

Bakeman R, Gottman JM. 1997. Observing Interaction: An Introduction to Sequential Analysis. Cambridge University Press, Cambridge, UK.

Black N. 1996. Why we need observational studies to evaluate the effectiveness of health care. *BMJ*, 312(7040): 1215-1218.

Boehm-Davis DA, Remington R. 2009. Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accid Anal Prev*, 41(5): 1124-1129.

Brixey JJ, Robinson DJ, Johnson CW, Johnson TR, Turley JP, Zhang J. 2007. A concept analysis of the phenomenon interruption. *Adv Nurs Sci*, 30(1): E26.

Brown BM, Dunsmuir WTM. 2010. Adjusting for length biased sampling when assessing the impact of interruptions on total time on task in a hospital emergency department setting. *Qual Saf Health Care*, 19(4): 284-289 web only supplement.

Chisholm CD, Dornfeld AM, Nelson DR, Cordell WH. 2001. Work interrupted: a comparison of workplace interruptions in emergency departments and primary care offices. *Ann Emerg Med*, 38(2): 146-151.

Chorney JM, Garcia AM, Berlin KS, Bakeman R, Kain ZN. 2010. Time-window sequential analysis: An introduction for pediatric psychologists. *J Pediatr Psychol*, 35(10): 1061-1070.

Cohen J. 1960. A coefficient of agreement for nominal scales. *Educ Psychol Meas*, 20(1): 37-46.

Coiera E. 2012. The science of interruption. *BMJ Qual Saf*, 21(5): 357-360.

Colligan L, Bass EJ. 2012. Interruption handling strategies during paediatric medication administration. *BMJ Qual Saf*, 21(11): 912-917.

Cook TD, Campbell DT. 1979. Quasi-experimentation: design & analysis issues for field settings. Houghton Miffin, Boston, MA, USA.

Czerwinski M, Horvitz E, Wilhite S. 2004. A diary study of task switching and interruptions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI 2004. ACM, Vienna, Austria, pp. 175-182.

Diggle PJ, Heagerty P, Liang KY, Zeger SL. 2002. Analysis of logitudinal data, 2nd ed. Oxford University Press, Oxford UK.

Finkler SA, Knickman JR, Hendrickson G, Lipkin M, Thompson WG. 1993. A comparison of work-sampling and time-and-motion techniques for studies in health services research. *Health Serv Res*, 28(5): 577-597.

Flynn EA, Barker KN, Gibson JT, Pearson RE, Berger BA, Smith L. 1999. Impact of interruptions and distractions on dispensing errors in an ambulatory care pharmacy. *Am J Health Syst Pharm*, 56(13): 1319-1325.

Gonzalez VM, Mark GJ. 2004. Constant, Constant, Multi-tasking Craziness: Managing Multiple Working Spheres, CHI2004. ACM, Vienna, Austria.

Greenland S. 1989. Modeling and Variable Selection in Epidemiologic Analysis. *Am J Public Health*, 79(3): 340-349.

Greenland S, Rothman KJ, Lash TL. 2008. Measures of effect and measures of association. In: Rothman KJ, Greenland S, Lash TL (Eds.), Modern Epidemiology. Lippincott Williams & Wilkins, Philadelphia, USA.

Grundgeiger T, Sanderson P. 2009. Interruptions in healthcare: theoretical views. *Int J Med Inform*, 78(5): 293-307.

Grundgeiger T, Sanderson P, MacDougall HG, Venkatesh B. 2010. Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *J Exp Psychol Appl*, 16(4): 317-344.

Gunal MM, Pidd M. 2006. Understanding accident and emergency department performance using simulation. In: Perrone LF, Wieland FP, Liu J, Lawson BG, Nicol DM, Fujimoto RM (Eds.), Winter Simulation Conference, Monterey, CA, USA.

Harr R, Kaptelinin V. 2007. Unpacking the Social Dimension of External Interruptions, GROUP '07. ACM, Florida, USA, pp. 399-408.

Herzog TN, Scheuren FJ, Winkler WE. 2007. Data quality and record linkage techniques. Springer, New York, USA.

Hill AB. 1965. The Environment and Disease: Association or Causation?, Proceedings of the Royal Society of Medicine, pp. 295-300.

Hintze JM, Volpe RJ, Shapiro ES. 2002. Best practices in the systematic observation of student behaviour. In: Thomas A, Grimes J (Eds.), Best practices in school psychology. National Association of School Psychologists, Washington DC, USA, pp. 933-1006.

Hosmer DW, Lemeshow S. 1999. Applied Survival Analysis. John Wiley & Sons, New York, USA.

Hougaard P. 1995. Frailty models for survival data. *Lifetime Data Anal*, 1(3): 255-273.

Jones SRG. 1992. Was there a Hawthorne effect? *Am J Sociol*, 98(3): 451-468.

Katidioti I, Taatgen NA. 2014. Choice in Multitasking: How Delays in the Primary Task Turn a Rational Into an Irrational Multitasker. *Hum Factors*, 56(4): 728-736.

Latorella KA. 1999. Investigating interruptions: Implications for flightdeck performance. National Aeronautics and Space Administration, Langley Research Center, Langley, USA.

Lebiere C, Anderson JR, Bothell D. 2001. Multi-tasking and cognitive workload in an ACT-R model of a simplified air traffic control task, Tenth Conference on Computer Generated Forces and Behavioral Representation, Norfolk, VA, USA.

Levitt SD, List JA. 2011. Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. *Am Econ J: Appl Econ* 3(1): 224-238.

Li SYW, Magrabi F, Coiera E. 2012. A systematic review of the psychological literature on interruption and its patient safety implications. *J Am Med Inform Assoc*, 19(1): 6-12.

Lin DY, Psaty BM, Kronmal RA. 1998. Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies. *Biometrics*, 54(3): 948-963.

Liu D, Grundgeiger T, Sanderson PM, Jenkins SA, Leane TA. 2009. Interruptions and blood transfusion checks: lessons from the simulated operating room. *Anesth Analg*, 108(1): 219-222.

Lopetegui MA, Bai S, Yen P-Y, Lai AM, Embi PJ, Payne P. 2013. Inter-Observer Reliability Assessments in Time Motion Studies: The Foundation for Meaningful Clinical Workflow Analysis. AMIA Annual Symposium, Washington, DC, USA.

Lopetegui MA, Yen P-Y, Lai A, Jeffries J, Embi P, Payne P. 2014. Time Motion Studies in Healthcare: What are we talking about? *J Biomed Inf*, 49: 292-299.

Loukopoulos LD, Dismukes RK, Barshi I. 2001. Cockpit interruptions and distractions: A line observation study, 11th International Symposium on Aviation Psychology, Columbus, Ohio.

Loukopoulos LD, Dismukes RK, Barshi I. 2009. The multitasking myth: Handling complexity in real-world operations. Ashgate Publishing Co., Surrey UK.

Mache S, Vitzthum K, Klapp BF, Groneberg DA. 2012. Doctors' working conditions in emergency care units in Germany: a real-time assessment. *Emerg Med J*, 29(e1).

Mark G, Gudith D, Klocke U. 2008. The cost of interrupted work: more speed and stress, Computer-Human Interactions. ACM, Florence, Italy, pp. 107-110.

Mark GJ, Voida S, Cardello AV. 2012. "A Pace Not Dictated by Electrons": An Empirical Study of Work Without Email. In: Konstan JA. (Ed.), CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas.

McFarlane DC. 1997. Interruption of people in human-computer interaction: A general unifying definition of human interruption and taxonomy. NRL-FR-5510-97-9870. Office of naval research, Arlington, VA, USA.

Mintzberg H. 1970. Structured observation as a method to study managerial work. *J Manag Stud*, 7(1): 87-104.

Nugus, P., Braithwaite, J., 2010. The dynamic interaction of quality and efficiency in the emergency department: Squaring the circle? *Soc Sci Med* 70(4): 511-517.

Pollitt LC, Reece SE, Mideo N, Nussey DH, Colegrave N. 2012. The Problem of Auto-Correlation in Parasitology. *PLoS Pathog*, 8(4): e1002590.

Rothman KJ, Greenland S, Poole C, Lash TL. 2008. Causation and causal inference, In: Rothman KJ, Greenland S, Lash TL. (Eds.), Modern Epidemiology. Lippincott Williams & Wilkins, Philadelphia, USA.

Salvucci DD, Taatgen NA, Borst JP. 2009. Toward a Unified Theory of the Multitasking Continuum: From Concurrent Performance to Task Switching, Interruption, and Resumption. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems. ACM, Boston, MA, USA, pp. 1819-1828.

Sasangohar F, Donmez B, Trbovich P, Easty AC. 2012. Not All Interruptions are Created Equal: Positive Interruptions in Healthcare, Human Factors and Ergonomics Society Annual Meeting, pp. 824-828.

Schildcrout JS, Heagerty PJ. 2005. Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. *Biostatistics*, 6(4): 633-652.

Schisterman EF, Cole SR, Platt RW. 2009. Overadjustment Bias and Unnecessary Adjustment in Epidemiologic Studies. *Epidemiology*, 20(4): 488-495.

Shadish WR, Cook TD, Campbell DT. 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin Company, Boston, USA.

Sierhuis M, Clancey WJ, Van Hoof RJJ. 2007. Brahms: A multiagent modelling environment for simulating work processes and practices. *Int J Simul Process Model*, 3(3): 134-152.

Strayer DL, Drews FA. 2006. Multi-tasking in the automobile, in: Kramer AF, Weigmann DA, Kirlik A. (Eds.), Attention: From theory to practice. Oxford University Press, Oxford, UK. Su NM, Mark GJ, 2008. Communication Chains and Multitasking, CHI2008. ACM, Florence, Italy, pp. 83-92.

TimeCaT 3.9. 2013. TimeCaT Time Capture Tool (www.timecat.org). The Ohio State University.

Trafton JG, Altmann EM, Brock DP, Mintz FE. 2003. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal.

Int J Hum Comput St, 58(5): 583-603.

Walter SR, Li L, Dunsmuir WTM, Westbrook JI. 2014. Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours. *BMJ Qual Saf*, 23(3): 231-241.

Watson JM, Strayer DL. 2010. Supertaskers: Profiles in extraordinary multitasking ability. *Psychon Bull Rev*, 17(4): 479-485.

Weigl M, Hoffmann F, Mller A, Barth N, Angerer P. 2014. Hospital paediatricians' workflow interruptions, performance, and care quality: a unit-based controlled intervention. *Eur J Pediatr*, 173(5): 637-645.

Weigl M, Mller A, Vincent C, Angerer P, Sevdalis N. 2012. The association of workflow interruptions and hospital doctors' workload: a prospective observational study. *BMJ Qual Saf*, 21(5): 399-407.

Weigl M, Mller A, Zupanc A, Glaser J, Angerer P. 2011. Hospital doctors' workflow interruptions and activities: an observation study. *BMJ Qual Saf*, 20(6): 491.

Westbrook, J.I., 2014. Interruptions and multi-tasking: moving the research agenda in new directions. *BMJ Qual Saf*, 23: 877-879.

Westbrook JI, Ampt A. 2009. Design, application and testing of the Work Observation Method by Activity Timing (WOMBAT) to measure clinicians' patterns of work and communication. *Int J Med Inform*, 78: S25-S33.

Westbrook JI, Coiera E, Dunsmuir WTM, Brown BM, Kelk N, Paoloni R, Tran C. 2010a. The impact of interruptions on clinical task completion. *Qual Saf Health Care*, 19(4): 284-289.

Westbrook JI, Woods A, Rob MI, Dunsmuir W, Day RO. 2010b. Association of interruptions with an increased risk and severity of medication administration errors. *Arch Intern Med*, 170(8): 683-690.

Wickens CD, 2002. Multiple resources and performance prediction. *Theor Issues Ergon Sci*, 3(2): 159-177.

Chapter 3

Defining clinical work

3.1 Chapter background

The quantitative analysis of observations of a continuous human process, such as clinical work, requires a conceptual framework by which observed action can be translated into quantitative data. Imposing some kind of conceptual structure will always be a simplification of what is actually observed, but the simplifying model can be devised in a way to enable analyses that generate insights of practical use. As discussed in the previous chapter, the main issues with the terminology and definitions used by many authors in the field are a lack of precision in definitions, a lack of consistency in the way certain terms are used between different studies, and an overly narrow focus on only defining isolated aspects of the work process such as 'interruptions'.

This chapter is included to make clear the system of concepts, terms and definitions used in this thesis to address the issues listed above. This system is used in each of the papers in subsequent chapters, but a full description is included in this chapter as there was insufficient scope in any single paper to describe the whole schema. These terms and their definitions do not purport to be universal, but rather were developed to facilitate the analysis of observed clinical practice in a way that is relevant to the clinical context, that can be consistently operationalised and that allows alignment with existing definitions as much as possible. This chapter addresses the second thesis objective.

3.2 States and transitions

In chapter 2 (section 5.3) an outline was given on how the work process of an individual clinician can be conceptualised as a sequence of states. This idea forms the basis of the concept of clinical work used in this thesis and from which definitions are derived and subsequently applied. It also allows existing definitions to be described via a set of common terms. A figure similar to Figure 2 in the previous chapter is also included here (see Figure 1 below), although it has been modified to illustrate the conceptual schema described below. While not all possible transitions are studied in this thesis, we define all those that plausibly occur in clinical work for both completeness and also to support future study.

The concept of *states* underpins a range of statistical techniques, Markov models being one of the better known examples. Essentially a process that evolves over time can be assigned to a particular category at certain period of that process. As a simple example, at any one time a person's health status can be considered healthy, diseased or dead. This basic idea is applied to the clinical work process such that a clinician's activity at a given time can be assigned to a category from a set of possible categories. The clinician may remain in that state for a certain period of time, then when the nature of activity changes the current state is deemed to end and a new one begun. This is diagrammatically represented in the 'sequence of states' section of the figure below.

In many observational studies of clinical work, clinicians' tasks are categorised by the type of clinical activity, for example, direct care, documentation, professional communication, and so on. This is represented in the figure by generic categories A, B and C. States can also be defined by other aspects of work such as whether there is any concurrent multitasking. For task 5 in the example shown in the figure, the addition of task 6 in parallel results in a change of state at transition T_8 and the new state is defined by both task category and multitasking status. In the similar figure in the previous chapter, a task from category A immediately following an interruption (labelled A_*) was considered a different state than other category A tasks. This illustrates that the specific definitions of states depend on the study, but in general in this way the continuous timeline of a single clinician's work is divided up into adjoining non-overlapping intervals, each assigned to a particular state and having start and end points in time.

Transitions describe the reasons for changing from one state to the next, and this section provides a glossary of transitions covering the majority of those that occur when observing the work process of an individual clinician. A key concept for describing transitions is the *prompt*. In general terms this is considered any occurrence that has the potential to initiate a change in state prior to task completion. It has precedent in Trafton et al.'s (2003) 'alert for the secondary task' (see Figure 1 in the previous chapter), however, it is tailored to be more applicable to work process observed in non-experimental settings. Prompts can be externally generated where some event not initiated by the clinician urges some kind of response, such as being asked a question or receiving a phone call. Alternatively, prompts can be internally generated where a potential motivator to change the course of the work process comes from within the clinician, such as remembering an urgent pending task or if there is a change in the clinician's prioritised mental list of tasks. Three types of transition are defined as follows:

1. *Sequential task completion* is used to refer to the change between tasks when they are performed one at a time, such that a new task is started only when the preceding one is complete. This is what Bluedorn et al. (1992) describe as monchronicity. It also represents one extreme of the multitasking continuum as proposed by Salvucci et al. (2009). At this end of the continuum there is no switching, while at the other end switching at the cognitive level occurs at the scale of fractions of a second giving the appearance of parallel task performance. In the figure below, transitions T_1 and T_2 are examples of sequential task completion.

2. The term *task-switching* describes suspending a task prior to its completion then switching to some other task or sequence of tasks. Switching prior to task completion is the feature that distinguishes this type of transition from sequential task completion. We consider two reasons for a clinician to task-switch. The first is due to an external prompt and is referred to as an *externally prompted* task-switch. This is shown in the figure at transition T_3 where an external prompt arrives during task 3 (the primary task) and in response the clinician suspends task 3 and switches to task 4 (the secondary task). Many definitions of interruptions in the healthcare literature are based around this idea. Some authors use the term *interruption* to describe the prompt (Chisholm et al., 2001; Czerwinski et al., 2004; Mache et al., 2012), some apply it to the task following the prompt (Li et al., 2012), while others use it to describe the sequence: primary task, prompt, task-switch, completion of secondary task, resumption of primary task (Boehm-Davis and Remington, 2009; Weigl et al., 2011). This sequence derives from Trafton et al.'s (2003) definition (see figure 1 in chapter 2), but it is important to note that prompts may be followed by transitions other than task-switching (see for example transition T_8), and tasks suspended by task-switching are not always resumed immediately upon completion

of the secondary task.

In contrast, we distinguish *internally prompted* task-switching. This term describes a single transition of a particular type, however, in the psychological literature, a sequence of internally prompted task-switches between two or more tasks being managed concurrently but not necessarily overlapping is sometimes called interleaved multitasking (Adler and Benbunan-Fich, 2012) or sequential multitasking (Savucci et al., 2009). In the figure, transitions T_4, T_5 and T_7 represent internally prompted task-switches since tasks in progress at the time of each switch are incomplete, yet there is no external prompt.

Transitions T_6 and T_12 represent a particular type of task-switch and, consistent with Trafton et al.'s (2003) terminology, this is referred to as *resumption*. Specifically, this is defined as resuming a suspended task after completion of some intervening sequence of action that followed the initial task-switch.

3. The final class of transitions is based around concurrent multitasking, that is, the performance of two of more tasks simultaneously. This corresponds to one extreme of the multitasking continuum where mental switching between tasks is so frequent as to give the outward impression of dual task performance (Salvucci et al., 2009). Two transitions related to multitasking are defined: 1) increasing the number of tasks occurring in parallel, called a *multitasking increase* (transitions T_8 and T_{10}), and 2) decreasing the number of such tasks (T_9 and T_{11}), a *multitasking decrease*. In practice clinicians largely transition between working on either one or two tasks, however, this definition allows for transitions involving more than two tasks, which can occur on occasion depending on how task categories are defined. Both types of transition can be externally or internally prompted.

3.3 Definitions used in this thesis: The prompt-response process

Much of the interruptions research in healthcare has essentially focused on external prompts and externally prompted task-switching in clinical practice. In this thesis external prompts are also examined, however, two key extensions to their conceptualisation as part of clinical work are applied. First, it is important to note that task-switching is not the only transition possible following an external prompt. For example, the second prompt in the figure above is followed by a multitasking increase. Several authors have discussed a number of strategies that can be used by clinicians in response to external prompts (Colligan and Bass, 2012; Collins et

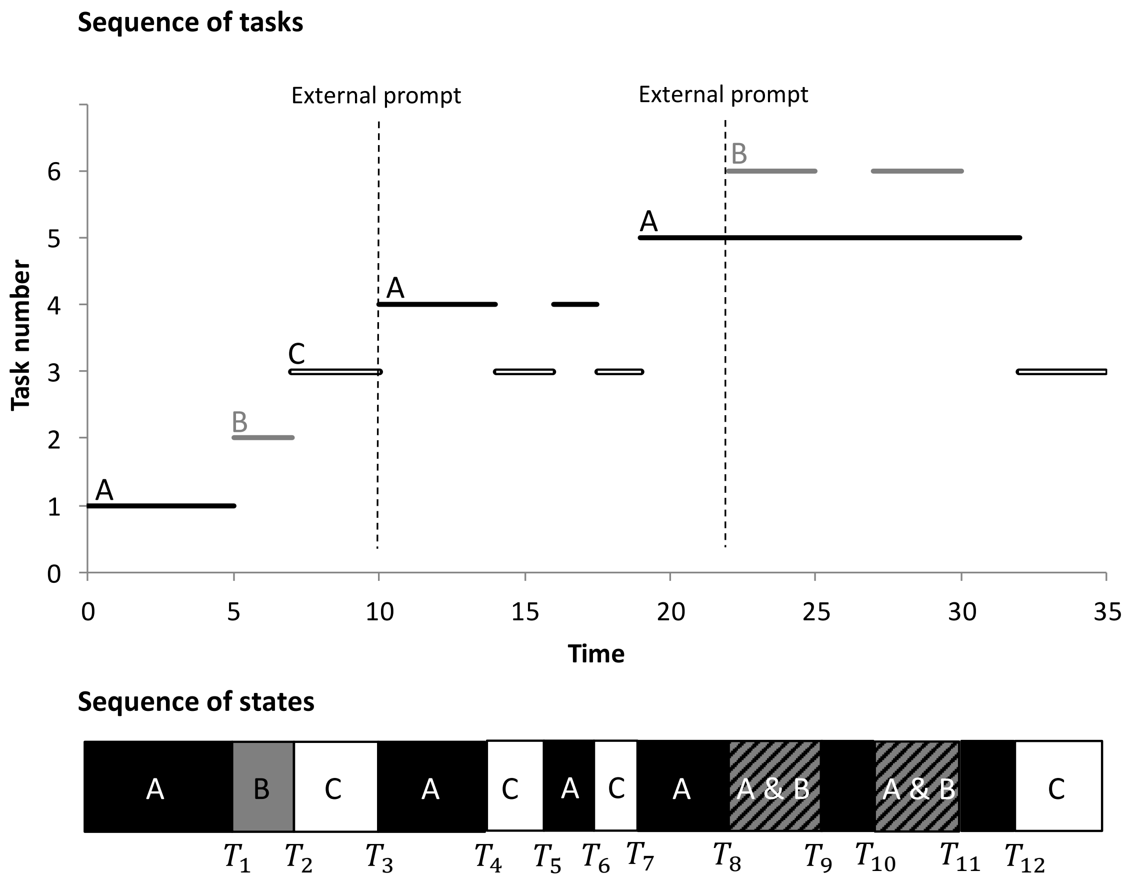


Figure 1. Illustration of the work process as a sequence of states and transitions.

al., 2007; Grundgeiger et al., 2010; Liu et al., 2009). Obviously task-switching is frequent in clinical work and has received much research attention, but a multitasking increase can also be initiated by such prompts. In the analyses in chapters 4 and 5 this is referred to as a multitasking strategy as the concepts of multitasking increase and decrease are superfluous to the scope of each individual article. Other strategies include *acknowledgement*, where certain prompts only require a simple nod or a one-word answer, *deferral* where requested action is deferred until a later time, and *deflection* which involves blocking, repelling or ignoring the prompt. In terms of transitions, these latter three strategies may involve switching or multitasking briefly to complete the acknowledgement, deferral or deflection, but it is more clinically relevant to distinguish them beyond states and transitions.

The second conceptual extension is to clearly distinguish the external prompt from the clinician's response to it. The conflation of these two ideas under the term 'interruption' in many studies is one of the main reasons for the ambiguity of that term. External prompts describe the actions of other actors (e.g. people, but also phones, alerts, etc.) directed towards the clinician. In contrast, the strategies described above represent the actions of the clinician. Thus the prompt-response process defines the interaction between clinicians and other actors. Treating prompts and clinicians' responses as separate but parallel streams of action is also fundamental to the statistical developments in chapter 6.

The analysis in the following chapter is the first application of the concept of the prompt response process. As this was early in the conceptual development, the term *trigger* was used to distinguish a subset of prompts that resulted in a change in the clinician's workflow, as opposed to prompts that we ignored. This was developed specifically for the data since there was no information on prompts, nor was it possible to separate internally and externally prompted multitasking.

References

Adler RF, Benbunan-Fich R. 2012. Juggling on a high wire: Multitasking effects on performance. *Int J Hum Comput St*, 70: 156-168.

Bluedorn AC, Felker Kaufman C, Lane PM. 1992. How many things do you like to do at once? An introduction to monochronic and polychronic time. *Acad Manage Exec*, 6(4): 17-26.

Boehm-Davis DA, Remington R. 2009. Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accid Anal Prev*, 41(5): 1124-1129.

Chisholm CD, Dornfeld AM, Nelson DR, Cordell WH. 2001. Work interrupted: a comparison of workplace interruptions in emergency departments and primary care offices. *Ann Emerg Med*, 38(2): 146-151.

Colligan L, Bass EJ. 2012. Interruption handling strategies during paediatric medication administration. *BMJ Qual Saf*, 21(11): 912-917.

Collins SA, Currie L,Patel VL, Bakken S, Cimino JJ. 2007. Multitasking by clinicians in the context of CPOE and CIS use. In: Congress Proceedings MedInfo 2007, Brisbane, Australia: 958-962.

Czerwinski M, Horvitz E, Wilhite S. 2004. A diary study of task switching and interruptions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI 2004. ACM. Vienna, Austria, pp. 175-182.

Grundgeiger T, Sanderson P, MacDougall HG, Venkatesh B. 2010. Interruption management in the intensive care unit: predicting resumption times and assessing distributed support. *J Exp Psych Appl*, 16(4): 317-334.

Li L, Hains I, Hordern T, Milliss D, Raper R, Westbrook J. 2015. What do ICU doctors do?: A multisite time and motion study of the clinical work patterns of registrars. *Crit Care Resus*, 17(3): 159-166.

Liu D, Grundgeiger T, Sanderson PM, Jenkins SA, Leane TA. 2009. Interruptions and blood transfusion checks: lessons from the simulated operating room. *Anesth Analg*, 108(1): 219-222.

Mache S, Vitzthum K, Klapp BF, Groneberg DA. 2012. Doctors' working conditions in emergency care units in Germany: a real-time assessment. *Emerg Med J*, 29(e1).

Salvucci DD, Taatgen NA, Borst JP. 2009. Toward a Unified Theory of the Multitasking Continuum: From Concurrent Performance to Task Switching, Interruption, and Resumption. In: Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems. ACM, Boston, MA, USA, pp. 1819-1828.

Trafton JG, Altmann EM, Brock DP, Mintz FE. 2003. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *Int J Hum Comput St*, 58(5): 583-603.

Weigl M, Muller A, Zupanc A, Glaser J, Angerer P. 2011. Hospital doctors' workflow interruptions and activities: an observational study. *BMJ Qual Saf*, 20: 491-497.

Chapter 4

The prompt-response process: a retrospective analysis

4.1 Chapter background

This paper represents the first application of the concept of the prompt-response process described in chapter 3 and was published in *BMJ Quality and Safety*. Existing datasets from three previous studies were used and their consistent observational methodology allowed comparison of effects, something rarely possible due to the general heterogeneity in definitions and methods in this field. Among authors who had previously discussed the concept of response strategies, none had quantitatively observed this in practice. Hence this article is an important first stage in applying an innovative and context-appropriate concept of the clinical work process. The paper was also fundamental in informing the design of the subsequent study of ED doctors in chapter 5.

Due to the lack of prior evidence on the prompt-response process, the analysis took an exploratory, rather than hypothesis testing, approach. Model building was used to select covariates significantly associated with clinicians' choice of task-switching versus multitasking, as opposed to including predetermined covariates in the models to test particular hypotheses. Variables were considered at several levels of the work system. Most of these were task-related variables such as type and duration of tasks, but others related to clinician characteristics, temporal factors and departmental workload. The supplementary (web only) table is included in this chapter following the main article.

To minimise the potential for bias due to the observational nature of the data, that is, to maximise internal validity, variation between (and correlation within) individuals was incorporated into the models through a random intercepts approach. The modelling also attempted to assess the influence of previous responses, i.e. autocorrelation, by including lag terms of the outcome variable as covariates. However, each lag term reduced the useable data in the model since the first row in each session (out of more than 900 sessions) had no prior value, resulting in a missing lag value. These aspects of the modelling approach align with points made in the paper in chapter 2 (section 3.4). As outlined in the introduction, these existing techniques are barely used in the study of clinical work and this analysis represents the novel application of those methods in this field.

This paper addresses the third thesis objective by implementing modelling techniques in a way not previously done in the study of clinical work. It also addresses the fourth objective by applying the prompt-response concept to existing data by providing new insights into factors at multiple levels of the work system that influence the interruptive aspects of clinical practice.

4.2 Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours

Author contributions

SRW conceptualised the study and all authors built on these concepts to design the study. SRW carried out the analysis while LL and WTMD provided input on statistical details of the analysis. SRW drafted the manuscript and led subsequent revisions. All authors contributed to the interpretation of results. LL, WTMD and JIW provided critical input into all versions of the manuscript.

Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours

Scott R Walter,[1] Ling Li,[1] William T M Dunsmuir,[2] Johanna I Westbrook[1]

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmjqs-2013-002097).

[1]Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, UNSW Medicine, University of New South Wales, Sydney, New South Wales, Australia
[2]Faculty of Science, School of Mathematics and Statistics, University of New South Wales, Sydney, New South Wales, Australia

Correspondence to
Scott R Walter, Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Level 1, AGSM Building, University of New South Wales, Kensington, NSW 2052, Australia; scott.walter@unsw.edu.au

Received 28 April 2013
Revised 18 August 2013
Accepted 12 September 2013
Published Online First 17 October 2013

To cite: Walter SR, Li L, Dunsmuir WTM, *et al*. *BMJ Qual Saf* 2014;**23**:231–241.

ABSTRACT

Objective To provide a detailed characterisation of clinicians' work management strategies.

Design 1002.3 h of observational data were derived from three previous studies conducted in a teaching hospital in Sydney, Australia, among emergency department (ED) doctors (n=40), ward doctors (n=57) and ward nurses (n=104). The rates of task-switching (pausing a task to handle an incoming task) and multitasking (adding a task in parallel to an existing task) were compared in each group. Random intercepts logistic regression was used to determine factors significantly associated with clinicians' use of task-switching over multitasking and to quantify variation between individual clinicians.

Results Task-switching rates were higher among ED doctors (6.0 per hour) than ward staff (2.2 and 1.8 per hour for doctors and nurses, respectively) and vice versa for multitasking rates (9.2 vs 17.3 and 14.1 per hour). Clinicians' strategy use was significantly related to the nature and complexity of work and to the person they were working with. In some settings, time of day, day of the week or previous chosen strategy affected a clinician's strategy. Independent of these factors, there was significant variation between individual clinicians in their use of strategies in a given situation (ED doctors p=0.04, ward staff p=0.03).

Conclusions Despite differences in factors associated with work management strategy use among ED doctors, ward doctors and ward nurses, clinicians in all settings appeared to prioritise certain types of tasks over others. Documentation was generally given low priority in all groups, while the arrival of direct care tasks tended to be treated with high priority. These findings suggest that considerations of safety may be implicit in task-switching and multitasking decisions. Although these strategies have been cast in a negative light, future research should consider their role in optimising competing quality and efficiency demands.

INTRODUCTION

The hospital work environment is complex with many competing demands on doctors' and nurses' time and cognitive capacity. Clinicians use various strategies to handle workload demands,[1] two of which we examine in detail: task-switching and multitasking. Both may be used in response to a stimulus, for example, being asked a question. A clinician can respond by pausing the current task to answer the question or continuing the task while concurrently answering the question, that is, he or she can multitask.

Task-switching is a strategy to deal with what is variously defined as an interruption. Interruptions are known to be an implicit part of clinical work.[2–7] Evidence of their cognitive burden has been documented in experimental psychology[8–10] and aviation,[11] where associations between interruptions and increased error,[9] stress[9 11] and task completion time[8–10] have been reported. Such outcomes in the clinical setting could have severe repercussions, and among studies examining interruptions to clinical work there are several reporting negative effects.[12–14]

Despite this evidence, the relationship between interruption and error in hospitals appears complex. Not all studies of this relationship found an association,[15] and several review papers discuss the overly simple assumption of a straightforward cause and effect pathway.[16 17] Interruptions can have a positive effect on patient safety and on efficiency.[16–18] Also, some studies suggest that clinicians use strategies like task-switching to deal with competing workload demands.[1 19] Indeed, there are many factors in the clinical environment that cause a broad spectrum of workflow discontinuities, from a simple question to a medical emergency, and the way in which clinicians manage these is not well understood. While multitasking has also been implicated as a source of cognitive load and a cause of error in other professions,[20 21] it too is central to clinical workload management, yet few studies have examined the role and impact of this strategy in detail[4 22] whether negative[23] or otherwise.

Thus, it is important to understand the role of task-switching and multitasking in workload management and what drives these work patterns. We aimed to provide a detailed characterisation of clinicians' work practices by addressing three main objectives: (1) to describe rates of task-switching and multitasking, (2) to identify factors that are associated with clinicians using one strategy over the other and (3) to examine to what extent strategy is determined by external as opposed to individual factors.

METHODS

Data sources

Data were derived from three studies conducted in a 400 bed metropolitan public teaching hospital in Sydney, Australia. The first dataset relates to observations of clinicians' work within the emergency department (ED) which had an average of 75 presentations per day. Participants comprised seven interns, 22 residents, six registrars and five physicians.[24] The other two datasets relate to doctors[6] and nurses[25] on four general medical and surgical wards (table 1). Of the ward doctors, 19 were interns, 20 residents and 18 registrars. Nurses comprised 19 registered nurse (RN) new graduates, 21 enrolled nurses, 13 year 2–4 RNs, 28 level 5+ RNs and 23 clinical nurse consultants and specialists. Both ward-based studies combine data from two periods of observation. Observation sessions occurred on weekdays between 07:00 and 19:00 h.

Data collection

Observations were carried out using the Work Observation Method By Activity Timing (WOMBAT).[26–28] This is a modified time and motion approach that collects information about work tasks, including task-switching and multitasking. The method involves observers shadowing participants while they are carrying out their normal work activities for 1–2 h at a time and recording every task performed in that time via a handheld computer. Information about multiple dimensions of work tasks is collected including, what task, with whom and using what information tools (eg, use of a phone, computer). All tasks are automatically time-stamped on data entry. If two or more tasks are observed to occur simultaneously they are recorded as multitasking, while if a task causes the premature cessation of another, this is recorded as task-switching. Figure 1 shows an exemplar screenshot of the WOMBAT data collection tool. Human research ethics approval was obtained from the study hospital and the University of New South Wales for all studies.

Operational definition of terms

The array of scenarios that sufficiently burden a clinician's cognition as to cause a change in work have been conceptualised in a number of ways.[29 30] Given that the term interruption has been muddied by inconsistent use in the literature,[31] we define an internally consistent set of terms which, as far as possible, are taken from the literature, but without use of the term interruption. A *primary task*[1 17 32–34] may be subject to some form of *trigger* that elicits a change in the work process; for example, while documenting, a phone rings. A clinician may respond to the trigger by replacing the primary task with a new, or *secondary*,[1 17 34] task (task-switching[35]); in our example, pausing documentation to answer the phone. Alternatively, a clinician may add a secondary task to the primary task (multitasking); for example, continuing to document while answering the phone. If the clinician ignores an attention-demanding prompt then the prompt is not considered a trigger as it does not result in workflow change. If action in response to a trigger is deferred via some short communication, this is still considered task-switching in itself. A trigger may be an externally or internally generated stimulus, but only internal triggers that result in multitasking have been recorded; for example, initiating a conversation while documenting. While in theory it may be

Table 1 Summary of data sources

Setting	Round	Date	Hours observed	Number of clinicians	Number of tasks	Number of wards	Number of observers
ED	–	July 2006–January 2007	210.8	40	8370	1	1
Ward doctors	1	July 2006–December 2006	150.9	19	9150	4	5
	2	March 2009–July 2009	205.4	38	8633	4	5
Ward nurses	1	July 2005–March 2006	251.1	52	14 318	4	8
	2	August 2008–December 2008	184.1	52	14 491	4	8

ED, emergency department.

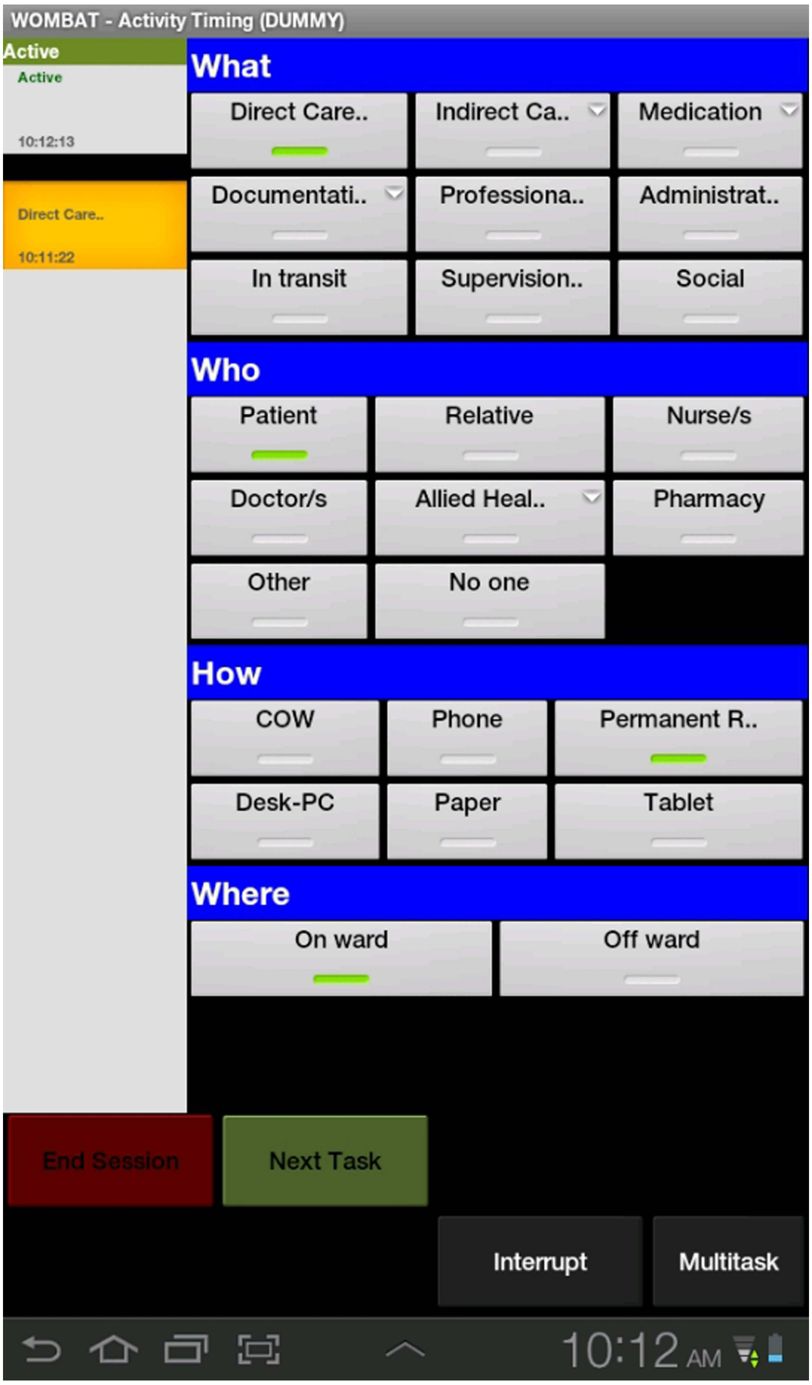


Figure 1 Work Observation Method By Activity Timing (WOMBAT) data collection tool.

possible to determine if a task-switch is due to an internal trigger, this is not captured in our data.

Statistical methods

Comparison of task-switching and multitasking rates

In order to address the first aim, rates of task-switching and multitasking per hour were compared. Rates with 95% CIs were calculated by primary task type using separate Poisson regression models for each group.[36] These modelled the rate of triggering events per unit time with trigger type (task-switch or multitask), task type and the interaction between the two as the covariates, and with time on task as the offset. If during a direct care task, for example, a clinician switched to a medication task, then the time spent on the direct care task was considered the exposure time for that trigger event and so contributed to the denominator of the direct care task-switching rate. Similarly, if an instance of multitasking was initiated during a direct care task, the duration of direct care contributed to the denominator of the direct care multitasking rate. Contrast statements were used to test the difference between task-switching rates and multitasking rates for each task type. The reporting of rates of task-switching and multitasking *per unit time* is common in the literature, but may not account for variation in the pace of work. Rates per task may be a reasonable alternative; however, defining the denominator for multitasking rates is non-trivial and rates of this form have not been included in this analysis.

Factors associated with clinicians' use of strategy

In line with the second aim, we used logistic regression to determine factors associated with clinicians' employment of task-switching (coded as 1) over multitasking (coded as 0) when triggered. The analysis used all instances of task-switching or multitasking, and included information on both the primary and secondary tasks associated with each trigger event. Random intercepts logistic regression was chosen to allow for correlation between outcomes within clinicians and wards, and to estimate clinician-specific rather than population-averaged effects. Potential covariates were: type and duration of primary task, tool involved in primary task (eg, computer), the type and duration of secondary task, tool and person involved in secondary task (eg, patient). Temporal factors included time of day, weekday and previous strategy used. Clinician seniority, years of experience and age were also potential predictors. The ED model considered the number of daily patient presentations as a proxy for clinical workload. A previous study of the ED data identified a temporal observer effect associated with the sole observer.[37] The two ward-based studies each had several observers. Hence, an observer effect was considered as a potential covariate in each model.

A model for each group was built by purposeful selection[38] with a significance level to retain covariates of 0.05. All models allowed for correlation within clinicians and the models for ward doctors and nurses also allowed for correlation within wards. Certain tasks may not be able to occur concurrently; however, the aggregation of tasks into task types (eg, direct care) means that all multitasking combinations between task types are able to occur. The definitions of each task type are given in online supplementary etable 1. These work task categories were developed through extensive piloting and consultation with clinicians[26 39] and have been shown to be valid in other settings.[28] Due to the challenge of interpreting interaction terms in a clinically useful way in this context, only first order terms were considered in the models.

Strategy determined by individual or external factors

The random intercepts models allow for variation between individuals due to unmeasured factors by allowing a different intercept value for each subject and we used this to assess the third aim of the study. The extent of this inter-clinician variation in strategy use is indicated by the variance between individual intercepts, after taking into account other factors in the model. This was done via a Wald test on the estimated intercept variance and its standard error. Significant intercept variance is evidence that there were other factors specific to individuals that influenced the use of strategy.

RESULTS

Among 8370 tasks observed in the ED, there were 1269 task-switches and 1942 instances of multitasking. For ward doctors, there were 795 task switches among 17 783 tasks in total, along with 6168 multitasking instances, and for ward nurses there were 28 809 tasks in total during which there were 800 task-switches and 4482 instances of multitasking (table 2).

Comparison of task-switching and multitasking rates

Overall, task-switching rates per hour were higher in the ED (6.0, 95% CI 5.7 to 6.4) than for doctors and nurses on wards (2.2, 95% CI 2.1 to 2.4 and 1.8, 95% CI 1.7 to 2.0). In contrast, multitasking rates per hour were lower in the ED (9.2, 95% CI 8.8 to 9.6 vs 17.3, 95% CI 16.9 to 17.8 and 14.1, 95% CI 13.8 to 14.5) (table 2). Rates for each task type also differed significantly for task-switching and multitasking rates within each group with the exception of ED rates of administrative tasks, answering a pager and supervision.

Factors associated with clinicians' use of strategy

The logistic regression models used to address the second aim identified several variables common to all work groups that were significantly associated with use of task-switching over a multitasking strategy. These were: the type of both primary and secondary tasks, the duration of the secondary task and whether a patient or a phone was involved in the secondary task (table 3). An observer effect of some kind was present in all models. In the ED model, the three time periods for the sole observer were significantly different, while in the other two models there were significant differences between observers. ORs refer to estimated fixed effects. Higher odds of task-switching (OR>1) is equivalent to lower odds of multitasking and vice versa.

Primary tasks

Clinicians in all work groups were most likely to switch tasks when triggered if the primary task involved documentation (table 3). Nurses were more likely to stop what they were doing during social

Table 2 Rates of task-switching and multitasking per hour by work group and primary task type

Work group	Task type	All tasks N	Task-switches N	Rate	95% CI	Multitasks N	Rate	95% CI
ED doctors	Direct care	992	204	3.4*	3.0 to 3.9	419	7.0	6.3 to 7.7
	Administrative	102	20	4.2	2.3 to 7.6	11	2.3	1.2 to 4.6
	Answering pager	20	2	8.2	1.8 to 38.0	3	12.3	3.6 to 42.2
	Combined medication tasks	447	56	5.3*	3.5 to 8.2	81	7.7	5.5 to 10.8
	Documentation	651	409	12.1*	8.9 to 16.5	314	9.3	7.3 to 11.9
	In transit	549	41	6.2*	3.8 to 9.9	64	9.6	6.7 to 13.7
	Indirect care	1928	397	7.2*	5.3 to 9.8	673	12.2	9.8 to 15.1
	Professional communication	3100	122	2.4*	1.7 to 3.4	335	6.6	5.2 to 8.4
	Social activities	528	10	0.8	0.4 to 1.8	31	2.6	1.6 to 4.1
	Supervisor/education	53	8	2.3*	1.0 to 5.3	11	3.1	1.6 to 6.2
	All tasks	**8370**	**1269**	**6.0***	**5.7 to 6.4**	**1942**	**9.2**	**8.8 to 9.6**
Ward doctors	Direct care	1459	82	1.2*	1.0 to 1.5	1035	14.9	14.0 to 15.8
	Administrative tasks	70	1	0.3*	<0.1 to 2.2	14	3.5	1.9 to 6.3
	Answering pager	256	8	2.4*	0.9 to 6.1	75	22.3	16.6 to 30.0
	Combined medication tasks	1996	46	1.7*	1.0 to 3.0	369	13.7	11.5 to 16.4
	Documentation	1746	164	3.0*	1.9 to 4.9	659	12.2	10.4 to 14.3
	In transit	2579	90	3.1*	1.9 to 5.2	460	15.9	13.4 to 18.9
	Indirect care	3356	247	3.1*	1.9 to 4.9	1709	21.1	18.4 to 24.2
	Professional communication	5747	114	0.9*	0.5 to 1.5	1616	12.8	11.1 to 14.7
	Social activities	416	12	0.3*	0.1 to 0.6	70	1.6	1.2 to 2.2
	Supervisor/education	158	31	1.1*	0.6 to 2.0	161	5.6	4.5 to 7.0
	All tasks	**17 783**	**795**	**2.2***	**2.1 to 2.4**	**6168**	**17.3**	**16.9 to 17.8**
Ward nurses	Direct care	4503	103	1.0*	0.8 to 1.2	801	7.4	6.9 to 8.0
	Ward related activities	412	18	1.3*	0.6 to 2.6	105	7.5	5.7 to 9.9
	Combined medication tasks	6462	244	3.1*	2.0 to 4.7	1097	13.9	11.8 to 16.3
	Documentation	1174	150	4.9*	3.1 to 7.6	356	11.6	9.5 to 14.0
	In transit	4006	54	1.8*	1.1 to 3.1	304	10.4	8.5 to 12.7
	Indirect care	4404	152	2.6*	1.6 to 4.0	826	14.0	11.8 to 16.5
	Professional communication	6796	55	0.6*	0.4 to 1.0	615	6.7	5.6 to 8.0
	Social	360	6	0.1*	0.0 to 0.3	16	0.3	0.2 to 0.5
	Supervision	485	16	1.1*	0.5 to 2.2	349	23.1	19.0 to 28.1
	Other	207	2	0.6*	0.1 to 3.0	19	5.7	3.4 to 9.6
	All tasks	**28 809**	**800**	**1.8***	**1.7 to 2.0**	**4482**	**14.1**	**13.8 to 14.5**

*Significantly different to multitasking rates for the same task type at the 0.05 level.
ED, emergency department.

activities compared with other tasks, but the opposite was true for both ED and ward doctors. The duration of primary task was only significant in the ward doctors' model where the odds of task-switching when triggered increased the longer a clinician spent on the primary task (OR 1.25, 95% CI 1.16 to 1.34) after adjusting for the effect of task type. ED clinicians were more likely to multitask if triggered while on the phone (OR 0.55, 95% CI 0.32 to 0.97).

Secondary tasks

Doctors in the ED and on wards most frequently suspended their primary task to perform direct care or answer a page when triggered, and nurses were most likely to switch tasks to perform direct care or for professional communication. For all groups, documentation and medication secondary tasks were more likely to be multitasked than most other task types. Longer duration of secondary tasks was significantly associated with increased likelihood of task-switching, and this effect was greater among nurses than both doctor groups (nurses OR 1.66, 95% CI 1.52 to 1.81; ED doctors OR 1.20, 95% CI 1.10 to 1.30; ward doctors OR 1.20, 95% CI 1.11 to 1.30).

People involved in the secondary task, in addition to the observed clinician, had a significant effect on a clinician's strategy. Clinicians in all groups were less likely to task-switch when triggered if a patient was involved in the secondary task (ORs between 0.21 and 0.47). ED doctors were more likely to switch tasks when triggered if a nurse or doctor was involved in the secondary task (OR 1.37, 95% CI 1.04 to 1.81; OR 1.54, 95% CI 1.19 to 1.99). Among ward doctors, a secondary task involving another doctor meant a reduced likelihood of stopping their primary task (OR 0.58, 95% CI 0.46 to 0.71). Similarly among nurses, secondary tasks with another nurse were associated with reduced odds of task-switching (OR 0.54, 95% CI 0.44 to 0.66).

Temporal factors

Among ED clinicians and ward nurses, strategy type was strongly associated with the strategy used for the previous trigger. Specifically, a clinician was more likely to switch tasks if the previous strategy was also a task-switch (ED OR 1.92, 95% CI 1.60 to 2.31; nurses OR 1.81, 95% CI 1.47 to 2.22). Time of day was a significant factor for ward doctors where the

Table 3 Factors significantly associated with the use of a task-switching strategy over a multitasking strategy when triggered

	ED doctors				Ward doctors				Ward nurses			
Fixed effect	**Number of tasks**	**OR**	**95% CI**	**p Value**	**Number of tasks**	**OR**	**95% CI**	**p Value**	**Number of tasks**	**OR**	**95% CI**	**p Value**
Primary task												
Task type				<0.001*				<0.001*				<0.001*
Direct care (ref)	623	1			1117	1			904	1		
Administrative/ward related	31	0.83	0.32 to 2.15	0.70	15	0.57	0.06 to 5.01	0.61	123	1.03	0.57 to 1.86	0.93
Answering pager	5	1.05	0.17 to 6.63	0.96	83	0.91	0.37 to 2.22	0.84		NA		
Combined medication tasks	137	0.94	0.58 to 1.50	0.78	415	1.17	0.72 to 1.91	0.53	1341	1.52	1.14 to 2.03	0.005
Documentation	723	1.55	1.10 to 2.19	0.01	823	1.52	1.02 to 2.27	0.04	506	1.72	1.20 to 2.45	0.003
In transit	105	0.56	0.33 to 0.94	0.03	550	1.47	0.96 to 2.26	0.08	358	1.28	0.87 to 1.89	0.22
Indirect care	1070	0.84	0.62 to 1.13	0.24	1956	0.87	0.60 to 1.25	0.45	978	1.02	0.76 to 1.38	0.89
Professional communication	457	0.70	0.49 to 1.00	0.05	1730	0.83	0.55 to 1.25	0.37	670	0.59	0.40 to 0.86	0.01
Social activities	41	0.47	0.20 to 1.10	0.08	82	0.64	0.27 to 1.48	0.29	22	1.62	0.45 to 5.74	0.46
Supervisor/education	19	0.86	0.27 to 2.78	0.80	192	0.85	0.47 to 1.54	0.59	365	0.37	0.20 to 0.67	0.001
Other		NA				NA			21	0.49	0.10 to 2.33	0.37
Tool involved†												
Permanent record	1108	NS			3096	NS			1770	1.36	1.09 to 1.69	0.01
Phone	123	0.55	0.32 to 0.97	0.04	370	NS			113	NS		
Nothing	1465	0.65	0.51 to 0.83	0.001	1917	0.72	0.57 to 0.91	0.01	2877	NS		
Task duration		NS				1.25	1.16 to 1.34	<0.0001		NS		
Secondary task												
Task type				<0.001*				<0.001*				<0.001*
Direct care (ref)	105	1			156	1			435	1		
Administrative/ward related	20	0.59	0.17 to 2.04	0.40	19	0.28	0.03 to 2.27	0.23	41	0.71	0.25 to 2.01	0.51
Answering pager	12	2.07	0.51 to 8.40	0.31	214	18.44	9.13 to 37.28	<0.0001		NA		
Combined medication tasks	216	0.23	0.12 to 0.45	<0.0001	1017	0.17	0.08 to 0.34	<0.0001	572	0.50	0.32 to 0.78	0.002
Documentation	151	0.15	0.06 to 0.36	<0.0001	1191	0.17	0.08 to 0.34	<0.0001	221	0.19	0.09 to 0.40	<0.0001
In transit	167	0.58	0.29 to 1.17	0.13	199	0.27	0.11 to 0.66	0.004	113	0.48	0.19 to 1.22	0.12
Indirect care	497	0.70	0.37 to 1.32	0.27	678	0.15	0.07 to 0.31	<0.0001	253	0.46	0.25 to 0.85	0.01
Professional communication	1761	0.56	0.30 to 1.04	0.07	3230	0.65	0.35 to 1.20	0.17	3329	0.85	0.59 to 1.22	0.38
Social activities	261	0.53	0.27 to 1.04	0.07	193	0.31	0.14 to 0.69	0.00	54	0.40	0.15 to 1.05	0.06

Continued

Walter SR, *et al*. *BMJ Qual Saf* 2014;**23**:231–241. doi:10.1136/bmjqs-2013-002097

Table 3 Continued

	ED doctors				Ward doctors				Ward nurses			
Fixed effect	**Number of tasks**	**OR**	**95% CI**	**p Value**	**Number of tasks**	**OR**	**95% CI**	**p Value**	**Number of tasks**	**OR**	**95% CI**	**p Value**
Supervisor/education	21	0.28	0.07 to 1.07	0.06	66	0.19	0.06 to 0.59	0.004	259	0.09	0.04 to 0.25	<0.0001
Other		NA				NA			11	2.79	0.67 to 11.55	0.16
Task duration		1.20	1.10 to 1.30	<0.0001		1.20	1.11 to 1.30	<0.0001		1.66	1.52 to 1.81	<0.0001
Person involved†												
Patient	216	0.34	0.20 to 0.56	<0.0001	1335	0.21	0.13 to 0.32	<0.0001	1045	0.47	0.35 to 0.64	<0.0001
Doctor	1152	1.54	1.19 to 1.99	0.001	4938	0.58	0.46 to 0.71	<0.0001	337	NS		
Nurse	697	1.37	1.04 to 1.81	0.02	880	NS			3758	0.54	0.44 to 0.66	<0.0001
No one	813	NS			742	0.58	0.38 to 0.90	0.01	271	0.46	0.26 to 0.82	0.01
Tool involved†												
Permanent record	48	3.54	1.42 to 8.84	0.01	2664	NS			917	0.72	0.55 to 0.95	0.02
Computer	187	0.48	0.28 to 0.83	0.01	-	NS			218	NS		
Phone	332	4.09	2.60 to 6.46	<0.0001	513	1.54	1.10 to 2.15	0.01	118	2.03	1.28 to 3.22	0.003
Nothing	2306	2.01	1.40 to 2.89	0.0002	2556	1.77	1.38 to 2.28	<0.0001	3935	NS		
Temporal factors												
Previous strategy‡												
Multitask (ref)	1862	1			5974	NS			4125	1.00		
Task-switch	1219	1.92	1.60 to 2.31	<0.0001	762				740	1.81	1.47 to 2.22	<0.0001
Time of day								0.01*				
<09:00	1031	NS			402	2.09	1.13 to 3.86	0.02	1264	NS		
09:00–11:59	598				2997	1.57	0.94 to 2.63	0.08	1580			
12:00–13:59	574				1109	1.50	0.87 to 2.57	0.14	799			
14:00–16:59	878				2143	2.04	1.22 to 3.40	0.01	1193			
≥17:00 (ref)	130				312	1			452			
Day of the week												0.04*
Monday	655	NS			1214	NS			951	1.15	0.83 to 1.58	0.41
Tuesday	333				1232				1595	0.76	0.57 to 1.02	0.07
Wednesday (ref)	802				1557				982	1		.
Thursday	791				1398				1137	0.87	0.66 to 1.16	0.35
Friday	630				1562				623	1.20	0.83 to 1.72	0.33

*p Value for global test for all categories together.
†Reference category is the complement; for example, reference category for 'patient' is 'not-patient'.
‡First task in each session classed as missing.
ED, emergency department;NA, variable not available; NS, variable not significant in multivariate model.

likelihood of task-switching was much lower at the end of the day (after 17:00) compared with all other times. Strategy use also varied by day of the week for nurses with task-switching more likely on Monday or Friday and less likely on Tuesday or Thursday. Seniority, age and clinician experience were not significant in any model.

Strategy determined by individual or external factors

For individual clinicians, the crude proportion of secondary tasks that were switched to ranged from less than 5% to over 80%. In each of the three models, after accounting for much of this variation through the model covariates, there was still significant variation between individuals according to the intercept variance assessed in relation to the third study aim (Wald test of intercept variance: $p=0.04$ for ED doctors and $p=0.03$ for ward staff). Inter-ward variation was zero for doctors and insignificant for nurses ($p=0.58$), indicating no evidence of difference in strategy use between wards within individual studies.

DISCUSSION

This study of over 1000 observation hours demonstrates that many factors affect clinicians' workload management strategies. ED doctors switched tasks frequently in contrast to a predominance of multitasking among doctors and nurses on wards. This high task-switching rate among ED doctors is consistent with other studies.[4 6 40 41] Increased task-switching frequency has been associated with work intensity[42 43] and with error in the clinical setting.[12 14] This suggests the elevated task-switching rate in EDs may be due to periods of higher work intensity which may increase error risk; although, the relationship among task-switching, work intensity and error risk is currently not well understood. In contrast, the rates of multitasking on wards were almost twice that of ED clinicians, suggesting that multitasking may be more easily carried out in settings of lower work intensity. This is consistent with the higher level of critical care provided in the ED compared with general wards. The inclusion of internally triggered multitasking may have increased multitasking rates and may partly explain why these rates are higher than overall task-switching rates in each work group.

The type of both primary and secondary task was a key factor in strategy use. In particular, documentation tended to be suspended in preference to most secondary tasks, but if documentation was the secondary task this was more likely to be performed in parallel with the primary task. Two studies of nurses' work report that the priority of secondary tasks relative to primary tasks is a determinant of the strategy used, with relatively higher priority secondary tasks necessitating task-switching.[1 19] An ED-based study also found that high priority triggers, such as calls to the patient resuscitation room, were more likely to cause task-switching than other less urgent triggers.[44] In light of this, documentation may be largely treated as lower priority compared with other task types. Under this line of reasoning, the fact that the primary task was more likely to be paused if the secondary task was direct care suggests direct care was generally treated as higher priority. These effects were seen in all work groups, and this may imply a degree of prioritising certain tasks over others.

The strong association between strategy use and task type may also be due to specific task combinations being naturally more suited to task-switching or multitasking strategies.[22] For example, while talking to a colleague, it is easier to simultaneously perform a routine manual task, but answering the phone would require task-switching. Further, particular tasks may not be able to occur concurrently (such as suturing two patients at once) and it is not known how the frequency of such mutually exclusive pairings may differentially affect the strategies that clinicians use.

The complexity, frequency and dissimilarity of secondary tasks can affect performance[10 45] and likely influence which strategy a clinician employs. Although secondary task duration is unknown at the time a trigger occurs, it was used as a proxy for cognitive demand assuming that on average longer tasks are more demanding. According to multiple resource theory, secondary tasks consuming only some cognitive processing resources can be multitasked, while those consuming all resources must be switched to.[17 46] Assuming then that secondary tasks causing task-switching are on average more demanding, then the observation that these tasks also tend to be longer suggests secondary task duration may indeed be a reasonable proxy for cognitive demand. However, without a direct measure of cognitive demand it is not possible to determine this with certainty. Although the number of daily presentations was not a significant factor in the ED, this measure may be too coarse to capture fluctuations in cognitive demand from one task to the next. The duration and complexity of secondary tasks have been associated with increased resumption lag, that is, time to reorientate when switching back to the primary task.[15 47] Also, the longer the primary task is suspended the more likely that it will not be resumed at the correct place in the sequence or that it will not be resumed at all.[16]

In the ED, doctors were less likely to pause a task for a patient than for another doctor or nurse. This may reflect the higher priority assigned to interactions with colleagues, or that such interactions are more cognitively demanding in the critical care setting. A similar response to patients was observed in wards, but there was also a lower likelihood of stopping work when triggered by a fellow clinician. This may indicate a different mode of inter-clinician interaction in non-critical care settings.

The results identified several temporal factors related to management of triggers that have not previously been reported. ED clinicians' and ward nurses' strategies were strongly associated with the strategy used for the previous trigger. This suggests the previous strategy may somehow influence the current one or that triggers of a similar cognitive demand arrive in clusters. Prompted by the significant change in strategy use with time of day among ward doctors, subsequent analysis identified that nurses' task-switching rates increased throughout the day, while ward doctors experienced a morning peak in multitasking rates, but a more constant rate of task-switching. Task-switching peaked in the ED in the late morning declining thereafter, while multitasking rates were steady. The significance of both previous strategy and time of day implies that sequencing and cumulative effects represent important factors in cognitive load and point the way for future investigation.

The significant intercept variance from the random intercept models suggests that there was individual variation in strategy use independent of external factors such as work tasks, people, tools and the hospital setting. This variation in individuals' strategy use was also not associated with age, experience or seniority, and is consistent with findings that cognitive characteristics and personality affect individual prospective memory strategies.[48] Addressing the third study aim, it appears that there are both external factors and factors related to the individual that determine the use of strategies. However, the observed inter-clinician variation cannot necessarily be attributed to differences in cognitive characteristics and may also be due to differing case mix or variation in specific tasks performed that are not captured by the relatively broad task categories. A study of paediatric nurses found that past experience with errors and the establishment of rigid routines are important clinician-specific factors that affect the way individuals manage triggers[1] and such factors may also have influenced inter-clinician differences in this study.

Under the framework that triggers are a potential cause of error, a number of frequency reduction and impact minimisation strategies have been suggested. These include system focused changes like ensuring information or resources are available when and where required,[2 16] and measures to prevent or minimise triggering during tasks of high working memory.[17] For the latter strategy, tasks with a high propensity to receive triggers and with high risk of patient harm are obvious targets;[31] for example, the implementation of 'interruption free zones' for preparing medications.[49] Clinician-specific strategies that promote resilience to disruption have been proposed, including the use of resumption cues to ensure correct resumption of a task sequence, training staff to time triggers for task boundaries,[9 50] and training clinicians for dealing with triggers to engender a sense of control and reduce stress.[17] However, the use or effectiveness of such strategies is largely underexplored.[51]

Major strengths of this study include the large sample size and inclusion of data from different clinician groups. The fact that data were collected using the same methodology lends validity to inter-group comparisons. The significant observer effect was a limitation that was not anticipated given the high inter-rater reliability scores in each of the original studies based on univariate assessments. This indicates that a multivariate assessment of inter-rater reliability may be necessary to minimise any observer effects in future studies. In this study, the observer effects were mitigated by adjusting for these in the models. All observations were conducted in the same hospital, so the results may not necessarily reflect practices in all hospitals. The inclusion of internally generated triggers for multitasking but not task-switching was also a limitation. This is discussed above in relation to comparing task-switching and multitasking rates, while for the logistic models some unmeasured bias may exist if the proportion of multitasks that are internally triggered is dependent on other factors. The data do not capture all the complexity of the human interactions under study, but these unmeasured factors were largely accounted for by the random intercepts modelling approach.

CONCLUSIONS

We have conducted a large sample, multi-setting, observational study of the way clinicians manage their work. The use of task-switching and multitasking in response to triggers was associated with many factors related to the work tasks, the people and tools involved in those tasks, as well as factors related to individual clinicians. Despite clear differences in strategy use among ED doctors, ward doctors and ward nurses, clinicians in all settings appeared to prioritise certain types of tasks over others. The low priority given to documentation was consistent in all groups, as was the high priority given to direct care secondary tasks. These findings suggest that considerations of safety and workload may be implicit in task-switching and multitasking decisions. Although these strategies have been cast in a negative light, future research should explore their role in optimising competing quality and efficiency demands.

Contributors SRW had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. SRW contributed to the conception of the study and carried out the analysis. All authors contributed to the study design, interpretation of results, drafting of the manuscript and have approved the final manuscript for submission.

Funding This study was funded by ARC Discovery grant DP110100090 and NHMRC Program grant 568612. SRW is funded by a NHMRC Postgraduate Scholarship.

Competing interests None.

Ethics approval University of New South Wales Human Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Colligan L, Bass EJ. Interruption handling strategies during paediatric medication administration. *BMJ Qual Saf* 2012;21:912–17.
2 Coiera E, Tombs V. Communication behaviours in a hospital setting: an observational study. *BMJ* 1998;316:673–6.
3 Alvarez G, Coiera E. Interruptive communication patterns in the intensive care unit ward round. *Int J Med Inform* 2005;74:791–6.
4 Chisholm CD, Collison EK, Nelson DR, *et al*. Emergency department workplace interruptions: Are emergency physicians "interrupt-driven" and "multitasking"? *Acad Emerg Med* 2000;7:1239–43.
5 Laxmisan A, Hakimzada F, Sayan OR, *et al*. The multitasking clinician: decision-making and cognitive demand during and after team handoffs in emergency care. *Int J Med Inform* 2007;76:801–11.
6 Westbrook JI, Ampt A, Kearney L, *et al*. All in a day's work: an observational study to quantify how and with whom doctors on hospital wards spend their time. *Med J Aust* 2008;188:506–9.
7 Mache S, Vitzthum K, Klapp BF, *et al*. Doctors' working conditions in emergency care units in Germany: a real-time assessment. *Emerg Med J* 2012;29:e1.
8 Altmann EM, Trafton JG. Task interruption: resumption lag and the role of cues. In: *Proceedings of the 26th annual conference of the Cognitive Science Society*; 5–7 August 2004, Chicago, IL: Lawrence Earlbaum Associates, Inc., 2004:43–8.
9 Bailey BP, Konstan JA. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Comput Human Behav* 2006;22:685–708.
10 Speier C, Valacich JS, Vessey I. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences* 1999;30:337–60.
11 Latorella KA. *Investigating interruptions: implications for flightdeck performance*. Hampton, VA: Langley Research Center, National Aeronautics and Space Administration, 1999.
12 Westbrook JI, Woods A, Rob MI, *et al*. Association of interruptions with an increased risk and severity of medication administration errors. *Arch Intern Med* 2010;170:683–90.
13 Parker J, Coiera E. Improving clinical communication. *JAMIA* 2000;7:453–61.
14 Flynn EA, Barker KN, Gibson JT, *et al*. Impact of interruptions and distractions on dispensing errors in an ambulatory care pharmacy. *Am J Health Sys Pharm* 1999;56:1319–25.
15 Magrabi F, Li SYW, Day RO, *et al*. Errors and electronic prescribing: a controlled laboratory study to examine task complexity and interruption effects. *JAMIA* 2010;17:575–83.
16 Grundgeiger T, Sanderson P. Interruptions in healthcare: theoretical views. *Int J Med Inform* 2009;78:293–307.
17 Li SYW, Magrabi F, Coiera E. A systematic review of the psychological literature on interruption and its patient safety implications. *JAMIA* 2012;19:6–12.
18 Rothschild JM, Landrigan CP, Cronin JW, *et al*. The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care. *Crit Care Med* 2005;33:1694–700.
19 Patterson ES, Ebright PR, Saleem JJ. Investigating stacking: How do registered nurses prioritize their activities in real-time? *Int J Ind Ergonom* 2011;41:389–93.
20 Horberry T, Anderson J, Regan MA, *et al*. Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accid Anal Prev* 2006;38:185–91.
21 Loukopoulos LD, Dismukes RK, Barshi I. Cockpit interruptions and distractions: A line observation study. In: *The 11th International Symposium on Aviation Psychology*. Columbus, OH: 2001.
22 Weigl M, Müller A, Zupanc A, *et al*. Participant observation of time allocation, direct patient contact and simultaneous activities in hospital physicians. *BMC Health Serv Res* 2009;9:110.
23 Collins S, Currie L, Patel V, *et al*. Multitasking by clinicians in the context of CPOE and CIS use. In: Kuhn K, Warren JR, Tze-Yun L. eds. *Congress Proceedings MedInfo 2007*; 22–24 August, Brisbane, Australia, 2007:958–62.
24 Westbrook JI, Coiera E, Dunsmuir WTM, *et al*. The impact of interruptions on clinical task completion. *Qual Saf Health Care* 2010;19:284–9.
25 Westbrook JI, Duffield C, Li L, *et al*. How much time do nurses have for patients? a longitudinal study quantifying hospital nurses' patterns of task time distribution and interactions with health professionals. *BMC Health Serv Res* 2011;11:319.
26 Westbrook JI, Ampt A. Design, application and testing of the Work Observation Method by Activity Timing (WOMBAT) to measure clinicians' patterns of work and communication. *Int J Med Inform* 2009;78:S25–33.
27 Westbrook J, Creswick NJ, Duffield CM, *et al*. Changes in nurses' work associated with computerised information systems: Opportunities for international comparitive studies using the revised Work Observation Method By Activity Timing (WOMBAT). In: Bakken S. ed. *11th International Congress on Nursing Informatics*; 23–27 June 2012; Montreal, Canada: AMIA, 2012.
28 Ballermann MA, Shaw NT, Mayes DC, *et al*. Validation of the Work Observation Method By Activity Timing (WOMBAT) method of conducting time-motion observations in critical care settings: an observational study. *BMC Med Inform Decis Mak* 2011;11:32.
29 Jett QR, George JM. Work interrupted: a closer look at the role of interruptions in organizational life. *Acad Manage Rev* 2003;28:494–509.
30 Brixey JJ, Robinson DJ, Johnson CW, *et al*. . A concept analysis of the phenomenon interruption. *Adv Nurs Sci* 2007;30: E26–42.
31 Coiera E. The science of interruption. *BMJ Qual Saf* 2012;21:357–60.
32 Biron AD, Loiselle CG, Lavoie-Tremblay M. Work interruptions and their contribution to medication administration errors: an evidence review. *Worldviews Evid Based Nurs* 2009;6:70–86.
33 Boehm-Davis DA, Remington R. Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accid Anal Prev* 2009;41:1124–9.
34 Weigl M, Hornung S, Glaser J, *et al*. Reduction of Hospital Physicians' Workflow Interruptions: A Controlled Unit-Based Intervention Study. *J Healthc Eng* 2012;3:605–20.
35 Czerwinski M, Horvitz E, Wilhite S. A diary study of task switching and interruptions. In: *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems 2004; 24–29 April, Vienna, Austria: ACM, 2004:175–82.
36 McCullagh P, Nelder JA. *Generalized linear models*. London: Chapman & Hall/CRC, 1989.
37 Brown BM, Dunsmuir WTM. Adjusting for length biased sampling when assessing the impact of interruptions on total time on task in a hospital emergency department setting. *Qual Saf Health Care* 2010;19:284–9 supplement. http://qualitysafety.bmj.com/content/suppl/2012/05/17/qshc.2009.039255.DC1/Technical_note_revised_Jan_2010.pdf
38 Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd edn. New York: John Wiley & Sons Inc., 2000.
39 Ampt A, Westbrook JI, Creswick NJ, *et al*. A comparison of self-reported and observational work sampling techniques for measuring time in nursing tasks. *J Health Serv Res Policy* 2007;12:18–24.
40 Chisholm CD, Dornfeld AM, Nelson DR, *et al*. Work interrupted: a comparison of workplace interruptions in emergency departments and primary care offices. *Ann Emerg Med* 2001;38:146–51.
41 Weigl M, Müller A, Zupanc A, *et al*. Hospital doctors' workflow interruptions and activities: an observation study. *BMJ Qual Saf* 2011;20:491–7.
42 Weigl M, Müller A, Vincent C, *et al*. The association of workflow interruptions and hospital doctors' workload: a prospective observational study. *BMJ Qual Saf* 2012;21:399–407.
43 Friedman S, Elinson R, Arenovich T. A study of emergency physician work and communication: a human factors approach. *Israeli J Emerg Med* 2005;5:35–42.
44 Allard J, Wyatt J, Bleakley A, *et al*. "Do you really need to ask me that now?": a self-audit of interruptions to the 'shop floor'practice of a UK consultant emergency physician. *Emerg Med J* 2012;29:872–6.
45 Mark G, Gudith D, Klocke U. The cost of interrupted work: more speed and stress. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; 5–10 April 2008, Florence, Italy. ACM, 2008:107–10.
46 Wickens CD. Multiple resources and performance prediction. *Theor Issues Ergon Sci* 2002;3:159–77.
47 Grundgeiger T, Sanderson P, MacDougall HG, *et al*. Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *J Exp Psychol Appl* 2010;16:317–34.
48 McDaniel MA, Einstein GO. Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Appl Cogn Psychol* 2000;14:S127–44.
49 Anthony K, Wiencek C, Bauer C, *et al*. No interruptions please: impact of a no interruption zone on medication safety in intensive care units. *Crit Care Nurse* 2010;30:21–9.
50 Monk CA, Boehm-Davis DA, Trafton JG. The attentional costs of interrupting task performance at various stages. In: *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*; 30 September—4 October 2002; Baltimore, MD, Santa Monica, CA: SAGE Publications, 2002:1824–8.
51 Raban MZ, Westbrook JI. Are interventions to reduce interruptions and errors during medication administration effective? A systematic review. *BMJ Qual Saf*. Published Online First: 26 Aug 2013. doi:10.1136/bmjqs-2013-002118

eTable 1. Definitions of task types for doctors and nurses.

Task type	Doctors	Nurses
Direct care	All tasks directly involved with patient care, includes direct communication with patient and/or family Indirect care	Tasks directly involved with patient care, e.g. direct communication with patient &/or family, bathing, applying dressings, nursing procedures etc.
Indirect care	All tasks indirectly related to patient care, including searching for scans or x-rays	All tasks indirectly related to patient care, eg reviewing results, planning care, washing hands, reviewing documentation, returning equipment
Medication tasks	All tasks associated with medication, includes prescribing, finding medication order, preparation, administration, documentation, discussion and clarification	All tasks associated with medication, includes preparation, administration, documentation, discussion & clarification
Documentation	Any recording of patient information on paper or computer including discharge summaries, but excluding medication related documentation	Documentation (paper and electronic), excludes medication documentation
Professional communication	All non-medication-related communication with another health professional, includes meetings, requests for medical consults and discussion around planning care	All non-medication related communication with another health professional includes ward & patient handover. Excludes medication related discussion
Administrative tasks	Any administrative activity that is not related to direct or indirect individual patient care, for example employment issues, bed allocations	N/A
In transit	Time between tasks and between patients	Time between tasks and between patients. Excludes movement between patients in a shared room and movement within a single room
Social	All non-work activity or communication, as well as tea and meal breaks	All non work communication, e.g. meal/tea breaks, personal calls
Supervision/education	Supervising others, including supervising students as well as attending education sessions	Supervising others, including students
Pager	Whenever the pager alerts, this is recorded as an interruption; only includes time taken to look at and turn off pager	N/A
Ward related activities	N/A	Ward activities, includes coordinating beds & staffing
Other	N/A	Any other task not included above

Chapter 5

The prompt-response process: a prospective study

5.1 Chapter background

The article in this chapter, published in *Applied Ergonomics*, builds on the study in the previous chapter by examining the relationship between prompts and clinicians' responses to those prompts using data from a study specifically designed around the concept of the prompt-response process. While the former study was an important first stage in applying this concept, the studies from which that data were derived did not record all potential strategies and did not separate prompts and responses. This motivated the design of a study based on the new work process conceptualisation using observational techniques consistent with the previous studies. As part of this, the variables, their categories and their operational definitions were tailored to capture a more complete record of the prompt-response interaction as well as key system factors: task-level aspects of the clinical work process, clinician attributes including psychometric measures, location within the department and measures of workload. As a result, this study generated a detailed picture of the ways in which the type and frequency of prompts varies with context. The study presents new evidence about a range of scenarios that represent significant differences in strategy use by clinicians dealing with prompts.

In addition to task observations recorded via the WOMBAT system, data on prescribing errors were collected by manual record review, time-specific departmental workload was derived from the patient tracking database, and individual workload was determined through the heart rate variability of observed participants. These multiple sources of data were collected to support a broader program of work. This paper uses a subset of those sources and is the first publication derived from them. The study was carried out in the emergency department of a large tertiary hospital and details of the design are included in the study protocol in appendix A.2.

The original aim for the analysis was to use a similar modelling approach as in the previous chapter, that is, a random intercepts model including autoregressive terms, but to extend the logistic form to a multinomial form to accommodate having five rather than two response categories (task-switching, multitasking, acknowledgement, deferral and deflection). Since this was the first study to apply the prompt-response concept in the design of a quantitative study, there was little prior information about relative frequencies of strategies and prompt types for the purpose of sample size calculation. This resulted in fewer than anticipated numbers in certain categories which limited the analytic possibilities relative to the original analysis plan. The complexity of a multilevel multinomial model with lag terms and the small counts in certain categories meant that there were issues with convergence, making the planned approach unfeasible. Hence, a similarly exploratory but more easily interpreted nonparametric model was used in the paper. The analysis represents a unique application of nonparametric regression to the study of clinical work and the model identified a range of scenarios in which the pattern of strategy use was significantly different.

The article addresses objective two of the thesis by applying the concept of the prompt-response process to a working context. It addresses objective three by employing an analytic technique not previously used in the study of clinical work, and in so doing it addresses aim four by generating new insights about the way doctors respond to various types of prompts and the way their responses vary depending on multiple factors.

5.2 Emergency doctors' strategies to manage competing workload demands in an interruptive environment: an observational workflow time study

Author contributions

The wider ED study was conceptualised by all authors, while the conceptualisation of this paper was led by SRW in collaboration with the other authors. SRW performed the analysis, interpreted the results, prepared the initial draft and led all revisions of the manuscript. All co-authors contributed to the interpretation of results and provided critical input on each version.

Emergency doctors' strategies to manage competing workload demands in an interruptive environment: an observational workflow time study

Scott R. Walter[a], Magdalena Z. Raban[a], William T.M. Dunsmuir[b], Heather E. Douglas[a] and Johanna I. Westbrook[a]

[a] *Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia*

[b] *School of Mathematics and Statistics, Faculty of Science, University of New South Wales, Sydney, Australia*

Abstract

An observational workflow time study was conducted involving doctors in the emergency department (ED) of a large Australian hospital. During 121.7 hours across 58 sessions, we observed interruptive events, conceptualised as prompts, and doctors' strategies to handle those prompts (task-switching, multitasking, acknowledgement, deferral and deflection) to assess the role of multiple work system factors influencing doctors' work in the ED. Prompt rates varied vastly between work scenarios, being highest during non-verbal solo tasks. The propensity to use certain strategies also differed with task type, prompt type and location within the department, although task-switching was by far the most frequent. Communicative prompts were important in patient treatment and workload management. Clinicians appear to adjust their communication strategies in response to contextual factors in order to deliver patient care. Risk due to the interruptive nature of ED communication is potentially outweighed by the positive effects of timely information transfer and advice provision.

1. Introduction

Hospital emergency departments (EDs) are known to be disruptive environments where doctors juggle many competing demands, workload is highly variable (Levin, 2006) and interruptive communication is commonplace (Coiera and Tombs, 1998). Such characteristics in a safety critical setting provide clear motivation to understand the complexities of clinical work and their implications for both patient safety and work efficiency. There has been considerable attention on the potential for error in healthcare settings, including EDs, due to interruptions and multitasking. However, the role of these aspects of clinical work in augmenting or mitigating error risk is still not well understood.

Of the extensive literature on interruption and multitasking in clinical work, two features stand out. First, many studies report highly aggregated and generally descriptive results. Many work system factors influence clinicians' work at multiple levels (Carayon et al., 2014; Werner and Holden, 2015) and assessing particular interrelationships while accounting for the many other potentially confounding factors is demanding for both study design and analysis (Walter et al., 2015). Studies to date have provided a solid foundation for understanding clinical work, but much of the complexity of healthcare settings is not captured. Deeper understanding requires more direct engagement with that complexity through refined conceptualisation of the work process and by designing studies that capture and examine the many factors affecting clinicians' work.

Second, most studies are motivated by the potentially negative effects of phenomena such as interruptions and multitasking. There is certainly evidence from experimental settings demonstrating negative outcomes (Altmann and Trafton, 2004; Bailey and Konstan, 2006; Speier et al., 1999), and there is also precedent in the aviation industry where fundamental changes in practice were implemented to minimise error related to interruption and multitasking (Latorella, 1999; Loukopoulos et al., 2009). However, healthcare settings are vastly different from cockpits or controlled computer-based experiments. Applying assumptions about negative effects may limit understanding by blinkering the research focus. Arguably it is first necessary to gain a sufficiently granular understanding of the functioning of clinical work to determine aspects that are effective and resilient and those that represent real but ameliorable error risk.

A common conceptualisation of an interruption assumes someone working on a primary task switches to a secondary task when prompted by some external event, before then resuming the primary task (Trafton et al., 2003). This underlies many definitions applied in healthcare studies, but is overly simple for such contexts (Werner and Holden, 2015; Walter et al., 2015). Interruption is often framed as a single entity, but rather it is an interaction between multiple actors where one prompts a response from the other. The prompt can take many forms (a question, a phone call, etc.), and the recipient has a range of possible strategies for responding to these external events. Such strategies include task-switching (analogous to Trafton et al.'s interruption concept above), multitasking (continuing the primary task while taking on a secondary task), deferral (delaying the secondary task), deflection (blocking the secondary task) and acknowledging (Colligan and Bass, 2012; Collins et al., 2007; Grundgeiger et al., 2010; Liu et al., 2009). Each response strategy can potentially introduce particular risks and benefits. For example, a task-switching strategy introduces the risk of forgetting to resume the suspended task, but may achieve the timely resolution of an urgent secondary task. The interaction between sources of prompts and clinicians' responses to them is the first step in the chain of events initiated by each prompt. Examining how clinicians apply different strategies in response to a broad spectrum of prompts is therefore a natural starting point for building understanding of the full impacts of such events.

Of the studies that applied some version of the concept of multiple response strategies, none have quantitatively examined ED doctors' work in detail. In this study we aimed to provide a detailed quantitative characterisation of the prompt-response process, including a multivariate analysis of the role of multiple contextual work system factors that influence doctors' work in the ED.

2. Methods

2.1 Setting and participants

This study was conducted in the acute section of an ED in a 440-bed metropolitan tertiary teaching hospital in Sydney, Australia, between July and October 2014. The ED operates 24 hours, receiving 140 presentations per day on average. The department comprises three sections: one for less serious (largely ambulatory, primary care) cases, a sub-acute section for stays of up to 24 hours for patient monitoring, and an acute section that deals with both acute and urgent cases. The 15 beds in the acute section, three of which are resuscitation beds, are situated around the periphery of the space and a central raised section is a staff-only area that forms the base of operations for doctors. ED doctors working as resident medical officers (RMO), senior resident medical officers (SRMO), registrars or consultants in the acute section at the time of the study were eligible to participate. RMOs have completed their medical degree and internship and have one to two years' clinical experience, and SRMOs have two to three years' experience. Registrars have at least three years' experience and are engaged in specialist emergency medicine training, while consultants have completed their specialist training. We refer to these classifications collectively as seniority. Interns were excluded from the study.

Potential participants were approached directly by researchers, provided with information about the study and invited to participate. Those that agreed to participate provided written informed consent. Of the 41 doctors approached, 36 (88%) agreed to participate. The study focused on clinical work during the day shift (0800 to 1800). The shift commences with a handover round from the nightshift team. At 1400 the evening shift staff arrive and the handover round between day and evening shift doctors occurs at around 1700.

2.2 Observational methods

We used a type of time and motion study known as a workflow time study (Lopetegui et al., 2014; Walter et al., 2015), where participants were shadowed by an observer continuously during each observation session ranging from 30 minutes to three hours in duration. During sessions the observer recorded time stamped information about each of the participant's tasks via a hand held tablet with the Work Observation Method by Activity Timing (WOMBAT) software (Westbrook et al., 2009). Observation sessions were distributed between 0800 and 1800 to capture the full spectrum of work done by doctors on day shift. As far as possible session times and participants were chosen to achieve approximately equal amounts of observed time between doctor seniority and across the time of day. Observations were carried out by two observers (the first and second authors) and several sessions of piloting were conducted before study commencement to ensure adequate inter-rater reliability (IRR). There were no significant differences between observers when comparing the proportions of tasks and proportions of time within categories for the main analysis variables. Additional assessments were conducted periodically during the study to ensure there was no significant observer effect.

In addition to direct observations, we also calculated a measure of workload using data from the patient load monitoring system combined with the observed number of doctors for each observation session. The metric is a modified version of Bernstein et al.'s EDWIN score (2003). This is calculated as $\sum n_i t_i / N$ where n_i is the number of patients in triage category i, t_i is the triage score (1=lowest, 5=highest) for that category, and N is the number of doctors working in the acute section. This was calculated specific to the time of each prompt and only for the acute section of the ED.

2.3 Definitions

In addition to time stamps, several variables were recorded for each observed task. Based on previous work (Westbrook et al., 2008; Westbrook et al., 2010), a classification of task types was developed in consultation with senior ED staff (Table 1) to capture doctors' activities in a clinically relevant way. The location of the doctor at the start of each new task was also recorded, as was the role of others involved in the task (e.g. other doctor, nurse, relative, etc.) and any tools used to perform the task (phone, computer, paper). The 'other' categories in Table 1 are a combination of several subcategories that were later collapsed due to small numbers in each of the subcategories.

In the broadest sense we conceptualised a prompt as an event (not necessarily anticipated by the recipient) that had the potential to elicit a response in the form of a change in workflow (Walter et al., 2014). This is comparable to what Trafton et al. (2003) describe as the alert for a secondary task, however, the term prompt is more adapted to non-experimental settings. Prompts can be internally or externally generated, but internal prompts are not observable. For the purpose of this study a prompt refers to an observable event relevant to the participant that urges some kind of response. There can be uncertainty about what observable events are considered prompts, so the prompt categories, along with operational definitions, were developed after piloting (Table 1). In a previous study, a trigger was defined as a subset of prompts that result in a change in workflow, assuming some prompts do not elicit a change, i.e. are ignored (Walter et al., 2014). We use the broader term prompt in preference to trigger as we aimed to capture instances of ignoring a prompt. Prompts were recorded as separate, albeit brief, tasks so that the characteristics of each prompt could be recorded (type, person, etc.) as distinct from characteristics of the preceding task or the subsequent response. Categories of prompt type and prompting person were combined both to create clinically meaningful categories and to ensure sufficient cell counts for analysis.

We also conceptualised a set of response strategies that clinicians use to deal with a prompt. Drawing on previous studies (Colligan and Bass, 2012; Collins et al., 2007; Grundgeiger et al., 2010; Liu et al., 2009), we defined five strategies that may be used in ED work and that were considered observable (Table 1). The first strategy is *task-switching* where, having received a prompt, the primary task is suspended and the task related to the prompt (secondary task) is addressed. This is comparable to what many studies of clinical work have defined as an interruption, although we prefer the term task-switch to avoid the definitional heterogeneity associated with interruption. Task-switching is also a term used in cognitive psychology to describe switching between two tasks occurring in parallel (Monsell, 2003). Our use is consistent with that definition, albeit a subset where the switching is externally prompted. Another strategy is *multitasking*, defined as continuing the primary task while concurrently taking on one or more other tasks related to the prompt. This is analogous to prompted dual task performance in the parlance of cognitive psychology (Pashler, 2000). Other strategies were *acknowledgement*, where certain prompts only required a simple nod or a one-word answer, *deferral* where requested action was deferred until a later time, and *deflection* which involved blocking, repelling or ignoring the prompt. Due to very low frequencies of deferral and deflection, these were subsequently combined into one category for the analysis. Both task-switching and multitasking were identified during the analysis phase. If the task after the prompt was different to the task preceding the prompt, this indicated task-switching. If the tasks before and after were the same but there was an additional task after the prompt, then this was coded as prompted multitasking. Deferral, deflection and acknowledgement could all be recorded as characteristics of the prompt.

2.4 Statistical methods

The final data consisted of one record for each prompt and included task-level, clinician-level and department-level information specific to each prompt. The task-level variables related to the prompted tasks (type of task in progress when the prompt arrived and whether this involved a computer or phone), the type of prompt, the strategy used in response to the prompt (task-switch, multitask, etc.), the location of the participant at the time of the prompt, and the time of day. The clinician-level variables comprised seniority, age and gender of the participants, and the department-level variable was the time-specific workload.

Prompt rates per hour were calculated by task type and according to where tasks were underway. For example, the prompt rate for ‘direct care at the bedside’ was calculated as the number of prompts arriving during direct care tasks occurring at the bedside, divided by the sum of the duration of all such tasks. Testing of differences between rates was done with univariate Poisson regression using PROC GENMOD in SAS 9.4.

A form of non-parametric regression - conditional inference trees - was used to identify scenarios where strategies were used in differing proportions (Hothorn et al., 2006). Briefly, this approach first partitions the data into two subsets with the split defined by a covariate value. The proportions of each strategy are compared between the two data subsets. This is repeated for all covariate values to find the most significant partition, as determined by a permutation test. The process continues within each partition to define further partitions until no more significant partitions can be found. Unlike other recursive binary partitioning methods, the conditional inference approach is not biased by the number of categories in each covariate and also avoids overfitting. This was implemented with the *ctree* function in the party package using R (Hothorn et al., 2006; R Core Team, 2014). The model was used to identify different patterns of response strategies according to task type, prompt type (including phone calls), response, location, time of day, whether the doctor was on the phone when prompted, clinician seniority, age and gender, and the time-specific departmental workload. Consultants carried a cordless phone specifically for communication with other hospital departments. All doctors could receive calls from fixed line phones that were located in the central area. All doctors also carried personal mobile phones but rarely took calls on them.

3. Results

3.1 Prompts

There were 36 participants in total comprising 10 consultants, 11 registrars, 9 SRMOs and 6 RMOs. We observed 965 prompts in 121.7 hours of observation. The majority of these prompts involved communication from either a doctor or nurse (31% and 36%, respectively) (Table 2). Sixty-one percent of prompts occurred in the central area of the acute section where much of the inter-clinician communication took place. Doctors were often prompted when performing documentation and computer-based indirect care tasks in the central area, representing 22% and 27% of all observed prompts. Many prompts also occurred when doctors were in transit (17%) between different areas within the ED.

There was an increase in the rate of prompts with doctors' seniority. Consultants received 10.6 prompts per hour on average, registrars 8.2, and just over 6 for SRMOs and RMOs (Table 2), and these differences were significant (type 3 test $p<0.001$). Prompt rates were significantly higher in the central area with 10.1 per hour, compared to 3.1 at the bedside (rate ratio [RR] 2.7, 95% CI 2.2, 3.4, $p<0.001$) and 7.0 in other areas of the ED (RR 1.5, 95% CI 1.3, 1.7, $p<0.001$). The highest prompt rate, when stratified by task type and location, occurred when doctors were in transit: 20.7 prompts per hour. This was strongly augmented by seniority, with consultants receiving 37.6 prompts per hour of time in transit compared to 22.8 for registrars and around 10 for residents (Table 2), and these differences were significant (type 3 test $p<0.001$). Rates of prompts during documentation and indirect care were the next highest, and the majority of these occurred in the central area (91% and 80%, respectively) where these tasks were largely computer-based.

3.2 Response strategies

The most frequent strategy used in response to prompts was task-switching, deployed for 68% of prompts (Table 3). Acknowledgement was the next most common (21%) and multitasking and deferral/deflection were relatively infrequent (8% and 3%, respectively). Of the variables considered in the classification model of response strategies, only primary task type, prompt type, location and whether or not the primary task involved a phone were significant in the recursive binary partitioning of strategies (Figure 1). Clinician level variables, time of day and departmental workload were not significant in the model. Whether the primary task involved a computer was also not significant, although since many documentation and indirect care tasks involved a computer there may have been some collinearity with task type.

Phones played a key role in strategy choice with doctors almost exclusively using task-switching (89% of strategies, see node 2 in Figure 1) when prompted by a phone call. If a doctor was on the phone when a prompt arrived they only opted for task-switching 46% of the time and were more likely to defer or deflect the prompt than in any other scenario (23%) (node 4). The remaining nodes related to prompts that did not involve a phone. Node 7 represents prompts occurring largely during direct and indirect care tasks outside the central area, and showed the highest use of multitasking compared to any other node (22%). Most of this multitasking occurred during indirect care either in the corridors or at the bedside. The scenario in node 7 also shows a relatively high proportion of acknowledgement (32%) and sixty percent of these instances were in response to communication from nurses. Nodes 8 and 10 represent over sixty percent of all prompts and show a similar profile of strategies, partly due to both being made up of prompts to computer-based tasks in the central area: indirect care for node 8 and documentation for node 10. Both show less than five percent deferral or deflection and around 20% acknowledgement. Node 11 has the highest proportion of acknowledgement (34%), and represents a mixture of professional communication prompts from staff other than doctors or nurses (e.g. administrative staff, allied health) and social prompts mainly from clinical staff.

4. Discussion

Our results provide the first detailed quantitative analysis of the ways doctors use strategies to respond to interruptive prompts and the extent to which their propensity for particular strategies changes with context. We have also provided a comprehensive account of the huge variation in the rate of prompts associated with several interacting contextual factors such as task type, seniority and location within the ED. This study also represents a unique application of a novel nonparametric modelling technique to generate new insights about clinical work.

A key finding was the extent to which we observed that the prompt-response process among clinicians was central to the way care was delivered to patients, as demonstrated by the predominance of prompts involving professional communication, largely in the central area of the ED. Residents and registrars managed multiple patients and prompted consultants regularly to seek advice. Similarly, the nurse manager frequently prompted consultants to discuss bed allocations in order to manage patient load within the ED. This practice meant junior doctors received regular advice and supervision, nurse managers were able to manage patient load, and consultants were able to keep abreast of all patients in the acute section. A corollary of this was frequent prompting of senior doctors, and the observed increase in prompt rate with increasing seniority reflects this advice-seeking system. Hence, consultants were observed to not only expect prompts but to actively encourage them, and the central area provided a hub for this communication to occur. This contrasts with the previous focus on the negative aspects of this communication practice (Coiera and Tombs, 1998; Chisholm et al., 2000; Woloshynowych et al., 2007). Consistent with the very high frequency of task-switching used in response to such prompts, consultants were observed to generally give their full attention and priority to these clinical discussions, and it was common to see other junior doctors or nurses waiting until discussions were over before prompting.

Prompts tended to arrive much more frequently when doctors were alone and engaged in non-verbal tasks such as indirect care, documentation or being in transit. Although there was some variation in prompt rates depending on location, prompt rates during these tasks were generally higher for all locations compared to other task types. Doctors engaged in solo non-communicative activities may be perceived as more 'interruptible', but such prompts may not necessarily have any less cognitive impact. Task-switching introduces risk of non-resumption or resuming at the wrong place in a task sequence (Grundgeiger and Sanderson, 2009). This could be particularly problematic during non-verbal tasks, for example if key information is omitted when documenting in patient notes, or if a doctor forgets to pass on vital information because they were prompted while in transit.

Although task-switching was the dominant strategy, there was an elevated frequency of multitasking during direct and indirect care outside the central area. Many of the prompted tasks were manual such as hand washing, gathering equipment, taking blood samples and sending samples for testing, and many of the prompts involved communication. To some extent this increased multitasking propensity can be explained by the fact that tasks of differing modalities, in this case manual and verbal, can more easily be performed in parallel than tasks of the same modality, e.g. verbal and verbal (Wickens, 2008). However, the cognitive load of multitasking may pose a risk for safety critical tasks such as preparing blood samples to send for testing. Although it has been suggested that workload may play a role in clinicians' propensity to multitask (Weigl et al., 2013), we found no evidence of that time-specific departmental workload was associated with strategy choice.

Much of the previous literature on interruptions in clinical work has involved relatively descriptive counting studies reporting interruptions rates (generally higher in EDs) and proportions of time spent on particular tasks (Coiera, 2012). To some extent these have formed the basis of blanket interventions aimed at reducing all interruptions (Raban and Westbrook, 2013). However, such approaches fail to appreciate the complexity of clinical work and the way in which prompts are integral to the provision of information required to deliver care. In seeking to go beyond aggregated descriptive analyses and by focusing on the ways in which competing tasks are managed in the ED, this study clearly suggests that generic interventions to reduce interruptions are unlikely to be successful or useful in the ED setting, and may even be detrimental to the provision of quality and timely care. Preventing interruptive events for particular individuals may result in issues for those working around them such as delayed information exchange or increased workload. The study of clinical work still has a long way to go to understand the complex interplay of interdependent factors that contribute to risks to safety in clinical practice.

One of the fundamental questions in the field is the extent to which prompts and response strategies translate into errors or reduced efficiency (Coiera, 2012). This is a challenging research question to study in health care, and very few of the vast number of studies in the field have attempted it, almost none in the ED context. Our results provide a springboard for further research by elucidating the first part of the causal pathway between prompts and errors, that is, the prompt-response interaction. The recording of information about prompts and responses separately, as well as capturing multiple dimensions associated with each factor, has identified factors that influence the rate at which doctors are prompted, their responses to those prompts and how these behaviours are integrated into everyday clinical practice of the ED. Examination of when and the types of tasks that clinicians multi-task has provided further nuanced information about how clinicians juggle competing priorities. Thus these factors can be taken into account in the design of future studies that aim to assess the impact of the nature of prompts and how they are handled on specific outcome and efficiency indicators. We have presented one approach for how this might be achieved (Raban et al., 2015).

Further areas of investigation could consider extending the amount of observation time per participant to enable a more fine-grained analysis of both individual differences and temporal factors. Investigation of the influence of physical layout of an ED on communication strategies would also be valuable, given that layout is a key component in the work system (Carayon et al., 2014), but there is little research investigating the link between layout and the disruptive nature of ED work. It is interesting to note that current Australian guidelines on ED design make no mention of this aspect of work (Australasian College for Emergency Medicine, 2014). Finally, simultaneous observation of an entire clinical team would allow tracking of information flow and assessment of resilience to prompt-related error at the team level.

4.1 Limitations

An ever present limitation of workflow time studies is the potential for observer presence to influence participant behaviour, often called the Hawthorne effect. Two recent studies examined this effect in clinical settings. One observed a significant effect of observer presence on hand washing compliance (Hagel et al., 2015), and the other reported an increase in radiologists' productivity when monitored (Kidwai and Abujudeh, 2015). Both studies involved overt performance scrutiny. In contrast, participants in this study were informed of the study aims which were arguably sufficiently ambiguous such that there was no 'right' way to behave. We also made it clear that performance was not being assessed and that all data would remain anonymous. During observation sessions we maintained some distance from the participant (approximately 2-4m), particularly during clinical procedures. Although we have no direct measure of observer influence, we made every effort to minimise it and saw no evidence that it occurred.

5. Conclusions

The complexity of clinical work in EDs is affirmed by the many factors we identified that influence the way doctors are prompted and how they respond. There is abundant experimental evidence of the potentially negative effects of frequent prompting and the need to constantly switch tasks or perform multiple things simultaneously. There are certainly scenarios more at risk of prompt-related errors than others and these are worthy of closer examination to determine targeted improvements to procedures or physical layout where appropriate. However, there is clearly also an important role for interruptive interaction in the ED, particularly for advice seeking and information transfer.

Our results indicate ways in which clinicians adjust their communication strategies in response to contextual factors in order to deliver patient care. The prompt-response process between consultants and junior doctors demonstrates an inbuilt safety net that appears to flex in response to clinical demands. Such understanding of the application of different strategies cannot be gained from experimental studies. Communication between clinicians is essential for quality patient care, and although there may be some risk (as yet under-explored) due to its interruptive nature, this is potentially outweighed by the positive effects of timely information transfer and advice provision. What appears a busy and disruptive environment may in fact be the workings of a resilient and adaptable system.

Funding: The work was supported by the Australian National Health and Medical Research Council [grant numbers 1055498, 1054146]; and the Australian Research Council [grant number DP160100943].

References

Altmann EM, Trafton JG. 2004. Task interruption: resumption lag and the role of cues. In: Proceedings of the 26th annual conference of the Cognitive Science Society; 57 August 2004, Chicago, IL: Lawrence Earlbaum Associates, Inc., pp. 43-48.

Australasian College for Emergency Medicine. 2014. Emergency department design guidelines. West Melbourne, Victoria, Australia: Document no. G15. URL: https://www.acem.org.au/getattachment/cde7e04a-fb7d-423a-bfef-217965809d7a/ Emergency-Department-Design.aspx. Accessed: 24 Nov 2015.

Bailey BP, Konstan JA. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Comput Human Behav*, 22: 685-708.

Bernstein SL, Verghese V, Leung W, Lunney AT, Perez I. 2003. Development and validation of a new index to measure emergency department crowding. *Acad Emerg Med*, 10: 938-942.

Carayon P, Wetterneck TB, Rivera-Rodriguez AJ, Schoofs Hundt A, Hoonakker P, Holden R, Gurses AP. 2014. Human factors systems approach to healthcare quality and patient safety. *Appl Ergon*, 45: 14-25.

Chisholm CD, Collison EK, Nelson DR, Cordell WR. 2000. Emergency department workplace interruptions: Are emergency physicians interrupt-driven and multitasking? *Acad Emerg Med*, 7: 1239-1243.

Coiera E. 2012. The science of interruption. *BMJ Qual Saf*, 21: 357-360.

Coiera E, Tombs V. 1998. Communication behaviour s in a hospital setting: an observational study. *Brit Med J*, 316: 673-676.

Colligan L, Bass EJ. 2012. Interruption handling strategies during paediatric medication administration. *BMJ Qual Saf.* 21: 912-917.

Collins S, Currie L, Patel V, Bakken S, Cimino JJ. 2007. Multitasking by clinicians in the context of CPOE and CIS use. In: Kuhn, K., Warren, J.R., Tze-Yun, L. eds. Congress Proceedings MedInfo 2007; 2224 August, Brisbane, Australia, pp. 958-962

Grundgeiger T, Sanderson P. 2009. Interruptions in healthcare: theoretical views. *Int J Med Inform*, 78: 293-307.

Grundgeiger T, Sanderson P, MacDougall HG, Venkatesh B. 2010. Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *J Exp Psychol Appl*, 16: 317-334.

Hagel S, Reischke J, Kesselmeier M, Winning J, Gastmeier P, Brunkhorst FM, Scherag A, Pletz MW. 2015. Quantifying the Hawthorne effect in hand hygiene compliance through comparing direct observation with automated hand hygiene monitoring. *Infect Cont Hosp Ep*, 36: 957-962.

Hothorn T, Hornik K, Zeileis A. 2006. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat*, 15: 651-674.

Kidwai AS, Abujudeh HH. 2015. Radiologist productivity increases with real-time monitoring: the Hawthorne effect. *J Am Coll Radiol*, 12: 1151-1154.

Latorella K. 1999. Investigating interruptions. Implications for flightdeck performance. Hampton, VA, USA: NASA Langley Research Center.

Levin S, France DJ, Hemphill R, Jones I, Chen KY, Rickard D, Makowski R, Aronsky D. 2006. Tracking workload in the emergency department. *Hum Factors*, 48: 526-539.

Liu D, Grundgeiger T, Sanderson P, Jenkins SA, Leane TA. 2009. Interruptions and blood transfusion checks: lessons from the simulated operating room. *Anesth Analg*, 108: 219-222.

Lopetegui M, Yen PY, Lai A, Jeffries J, Embi P, Payne P. 2014. Time motion studies in healthcare: what are we talking about? *J Biomed Inform*, 49: 292-299.

Loukopoulos LD, Dismukes RK, Barshi I. 2009. The multitasking myth: Handling complexity in real-world operations. Farnham: Ashgate.

Monsell S. 2003. Task switching. *Trends Cogn Sci*, 7: 134-140.

Pashler H. 2000. Task switching and multitask performance. In: Monsell S, Driver J, eds. Control of Cognitive Processes: Attention and Performance XVIII. USA: International Association for the Study of Attention and Performance, pp. 277-307.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
URL:http://www.R-project.org/.

Raban MZ,Westbrook JI. 2013. Are interventions to reduce interruptions and errors during medication administration effective? A systematic review. *BMJ Qual. Saf*, online first 26 Aug 2013, doi:10.1136/bmjqs-2013-002118.

Raban MZ, Walter SR, Douglas HE, Strumpman D, Mackenzie J, Westbrook JI. 2015. Measuring the relationship between interruptions, multitasking and prescribing errors in an emergency department: a study protocol. *BMJ Open*, 5. doi:10.1136/bmjopen-2015-009076.

Speier C, Valacich JS, Vessey I. 1999. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sci*, 30: 337-60.

Trafton JG, Altmann EM, Brock DP, Mintz FE. 2003. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *Int J Hum Comput St*, 58: 583-603.

Walter SR, Dunsmuir WTM, Westbrook JI. 2015. Studying interruptions and multitasking in situ: The untapped potential of quantitative observational studies. *Int J Hum Comput St*, 79: 118-125.

Walter SR, Li L, Dunsmuir WTM, Westbrook JI. 2014. Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours. *BMJ Qual Saf*, 23: 231-241.

Weigl M, Mller A, Sevdalis S, Angerer P. 2013. Relationships of multitasking, physicians' strain, and performance: an observational study in ward physicians. *J Patient Saf*, 9: 18-23.

Werner NE, Holden RJ. 2015. Interruptions in the wild: Development of a sociotechnical systems model of interruptions in the emergency department through a systematic review. *Appl Ergon*, 51: 244-254.

Westbrook JI, Ampt A. 2009. Design, application and testing of the Work Observation Method by Activity Timing (WOMBAT) to measure clinicians' patterns of work and communication. *Int J Med Inform*, 78: S25-33.

Westbrook JI, Ampt A, Kearney L, Rob MI. 2008. All in a day's work: an observational study to quantify how and with whom doctors on hospital wards spend their time. *Med J Aust*, 188: 506-509.

Westbrook JI, Coiera E, Dunsmuir WTM, Brown BM, Kelk N, Paoloni R, Cuong T. 2010. The impact of interruptions on clinical task completion. *Qual Saf Health*

Care, 19: 284-289.

Wickens CD. 2008. Multiple resources and mental workload. *Hum Factors*, 50: 449-455.

Woloshynowych M, Davis R, Brown R, Vincent C. 2007. Communication patterns in a UK emergency department. *Ann Emerg Med*, 50: 407-413.

Table 1. Category definitions for variables used in analysis

Variable	Categories	Details
Clinician role	Consultant	Completed specialist training and has at least 5 years of experience
	Registrar	Engaged in specialist training and has at least 3 years of experience
	Senior resident medical officer (SRMO)	Completed medical degree and has 2-3 years of experience
	Resident medical officer (RMO)	Completed medical degree and has 1-2 years of experience
Type of task in progress when prompted	Direct care	Any activity directly related to the care of a patient including taking patient history, admitting, examining or escorting a patient, performing or assisting with medical procedures, communicating with the patient or relative
	Indirect care	Any activity related to the care of a patient that doesn't involve direct interaction with the patient. This includes reviewing patient records, documents, imaging or test results, ordering tests, gathering and returning equipment for procedures, and washing hands
	Communication	Professional or social face to face communication with hospital staff
	Documentation	Any recording of patient-related information on paper or computer
	In transit	Moving between areas within the ED
	Other task	A catch-all category for less frequent tasks comprising prescribing medication, supervising or instructing students and taking breaks
Location within the ED	Central area	Raised area in centre of acute section where most computer terminals were located and formed the base of operations for doctors and the nurse manager
	Bedside	The area within the curtains surrounding each bed, including in the area containing the three resuscitation beds
	Other location	A catch-all category for any other location. Predominantly this relates to the corridor between the central area and the beds, but also includes the ambulance bay, x-ray room, store room or areas outside the acute section.
Prompt type	Professional communication from doctor	Communication from another doctor about patient care or departmental management
	Professional communication from nurse	Communication from a nurse about patient care or departmental management
	Social communication	Non work-related communication
	Phone	Phone call intended for the participant
	Other prompt	A catch-all category for less frequent prompts encompassing announcements or alarms over the PA system, equipment alarms
Response Strategy	Task-switching	Suspending the primary task to attend to a secondary task
	Multitasking	Continuing the primary task while also attending to the secondary task
	Acknowledging	Responding to certain prompts with a brief word or gesture of acknowledgement
	Deferral/Deflection	Delaying the secondary task until a later time or avoiding the need to deal with it by indicating unavailability, delegating it to someone else or ignoring it entirely

Table 2. Rates per hour (and counts) of prompts by original task type, location, doctor seniority and prompt type

Task in progress when prompted	Location	Seniority				Prompt type					TOTAL
		RMO	*SRMO*	*Registrar*	*Consultant*	*Professional communication with Doctor*	*Professional communication with Nurse*	*Social communication*	*Phone*	*Other Prompt*	
Professional communication	*Bedside*	0	0	5.0 (4)	4.6 (7)	1.0 (3)	2.1 (6)	0	0.3 (1)	0.3 (1)	3.8 (11)
	Central	3.8 (13)	4.0 (17)	2.4 (12)	5.4 (45)	0.8 (17)	1.7 (36)	0.1 (3)	0.9 (18)	0.6 (13)	4.1 (87)
	Other location	1.4 (3)	1.4 (4)	3.3 (9)	4.3 (20)	0.8 (10)	1 (13)	0.2 (3)	0.4 (5)	0.4 (5)	2.9 (36)
Direct care	*Bedside*	1.8 (6)	2.8 (11)	1.9 (10)	3.3 (17)	0.5 (9)	1.4 (24)	0	0.1 (2)	0.5 (9)	2.5 (44)
	Other location	2.7 (6)	1.7 (6)	4.5 (6)	5.4 (9)	0.1 (1)	1.7 (15)	0.3 (3)	0.6 (5)	0.3 (3)	3.0 (27)
Documentation	*Bedside*	0	0	10.1 (3)	33.4 (3)	2.0 (1)	5.9 (3)	0	0	3.9 (2)	11.8 (6)
	Central	11.2 (40)	11.2 (56)	12.4 (68)	19.0 (45)	4.1 (67)	4.0 (66)	2.3 (37)	1.2 (19)	1.2 (20)	12.7 (209)
	Other location	0	12.1 (4)	15.8 (4)	15.8 (7)	4.4 (5)	2.7 (3)	1.8 (2)	1.8 (2)	2.7 (3)	13.3 (15)
In transit	*Other location*	9.3 (15)	10.8 (22)	22.8 (49)	37.6 (77)	6.4 (50)	8.4 (66)	2.7 (21)	1.0 (8)	2.3 (18)	20.8 (163)
Indirect care	*Bedside*	44.1 (5)	4.4 (1)	18.7 (10)	21.8 (8)	5.7 (7)	4.0 (5)	0.8 (1)	0.8 (1)	8.1 (10)	19.4 (24)
	Central	16.3 (38)	12.7 (61)	17.6 (65)	26.6 (97)	7.8 (113)	5.0 (72)	1.8 (26)	1.9 (27)	1.6 (23)	18.0 (261)
	Other location	5.5 (6)	9.6 (10)	9.3 (12)	26.1 (14)	1.5 (6)	5.5 (22)	1.5 (6)	0.3 (1)	1.8 (7)	10.6 (42)
Other	*Bedside*	0	3.7 (1)	0	7.6 (2)	1.1 (1)	2.2 (2)	0	0	0	3.3 (3)
	Central	6.3 (5)	7.5 (11)	5.5 (7)	4.4 (4)	1.8 (8)	2.5 (11)	0	1.4 (6)	0.5 (2)	6.1 (27)
	Other location	2.5 (6)	1.2 (3)	1.0 (1)	0	0.4 (3)	0.5 (4)	0	0.1 (1)	0.3 (2)	1.3 (10)
TOTAL		6.1 (143)	6.3 (207)	8.2 (260)	10.6 (355)	2.5 (301)	2.9 (348)	0.8 (102)	0.8 (96)	1.0 (118)	7.9 (965)

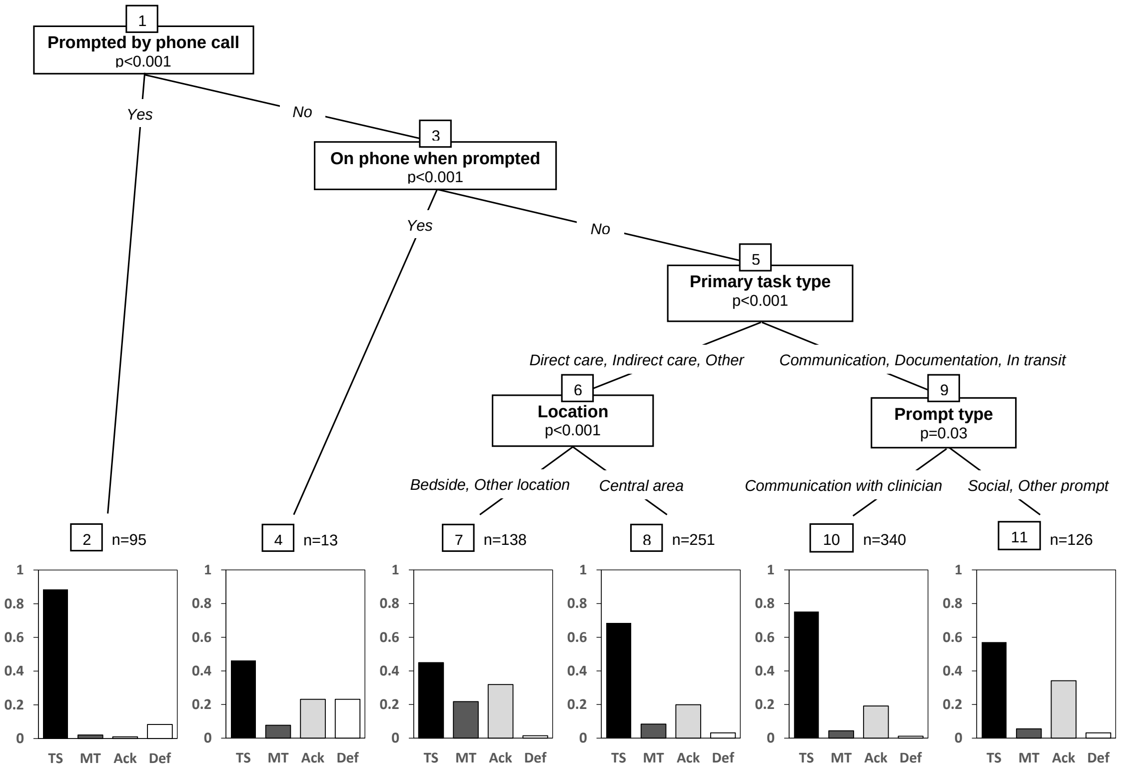
Note: RMO=resident medical officer, SRMO=senior resident medical officer

Table 3. Counts (and proportions) of strategies used in response to prompts, by doctor seniority and prompt type.

Response strategy	Seniority				Prompt type					TOTAL
	RMO	*SRMO*	*Registrar*	*Consultant*	*Professional communication with Doctor*	*Professional communication with Nurse*	*Social communication*	*Phone*	*Other Prompt*	
Task-switch	89 (62.2)	129 (62.3)	175 (67.3)	261 (73.5)	205 (68.0)	239 (68.7)	54 (52.9)	85 (89.5)	71 (60.2)	**654 (67.8)**
Multitask	14 (9.8)	20 (9.7)	15 (5.8)	27 (7.6)	22 (7.3)	31 (8.9)	13 (12.7)	2 (2.1)	8 (6.8)	**76 (7.9)**
Acknowledge	35 (24.5)	55 (26.6)	60 (23.1)	56 (15.8)	66 (22.0)	70 (20.1)	34 (33.3)	1 (1.0)	35 (29.7)	**206 (21.4)**
Defer/deflect	5 (3.5)	3 (1.4)	10 (3.8)	11 (3.1)	8 (2.7)	8 (2.3)	1 (1.0)	8 (8.4)	4 (3.4)	**29 (3.0)**
TOTAL	**143**	**207**	**260**	**355**	**301**	**348**	**102**	**96**	**118**	**965**

Note: RMO=resident medical officer, SRMO=senior resident medical officer

Figure 1. Proportions of response strategies used by ED doctors for subgroups identified by a conditional inference tree model.



Notes: TS=task-switching, MT=multitasking, Ack=acknowledging, Def=defer/deflect; p-values indicate significance of the difference between the two groups partitioned at a given node.

Chapter 6

The Poisson mixture model: theoretical details

6.1 Chapter background

The article in this chapter proposes new statistical methodology to address a fundamental question in the study of clinical work: how to assess the effect of task-switching, particularly externally prompted task-switching, on the time to complete tasks? A particular type of time cost (resumption lag) has been well documented in experimental studies, where the start and end of computer-based tasks are clearly defined and can be automatically recorded with software. In non-experimental settings such as hospitals, measuring these effects is very challenging. Since little is known about the efficiency implications of interruptive events, including task-switching, in healthcare contexts, there has been a clear need for methodology that assess the effect of task-switching on task completion time in these uncontrolled settings.

The proposed technique extends the only existing method in such a way as to make it more applicable to the heterogeneous data from observations of clinicians at work. It was developed over the majority of the candidacy period in collaboration with the authors of the original method. The challenge in developing methods of this type is shown by the fact that several promising early approaches that we tried proved inadequate and the final method went through many iterations before we had something that was theoretically sound, feasible to implement and improved on the previous technique. The new method has the potential to be applied to other events that occur frequently during clinicians' work, such as multitasking. In so doing it opens up a new way to understand the influence of interruptive events on clinical work.

This paper addresses the third objective of the thesis by developing new statistical methodology to address an important question specific to the observational study of clinical work. It also addresses the fourth objective by applying the method to data from observations of ED doctors.

6.2 Assessing the impact of task-switching on completion of clinical tasks in the presence of length bias

The manuscript has been submitted to *The Journal of the Royal Statistical Society: Series C (Applied Statistics)*.

Author contributions

SW proposed extending the Poisson assumption of the original method to more flexible assumption. All authors collaborated on the development of the statistical details related to that extension. SW performed the simulations, the example analysis and drafted the paper. BMB and WTMD provided critical input on each version of the paper.

Assessing the impact of task-switching on completion of clinical tasks in the presence of length bias

Scott R. Walter[a*], William T.M. Dunsmuir[b], B. M. Brown[b]

Clinical work is characterised by frequent interjection of events causing clinicians to switch from a primary task to deal with an incoming secondary task, before then returning to complete the primary task. Such task-switching is associated with several negative effects, including modification of the primary task completion time. An increase in task length due to task-switching implies reduced efficiency, while decreased length suggests hastening to compensate for the increased workload brought by the unexpected secondary tasks, which is a potential safety issue. Tasks that are naturally longer are more likely to have one or more task-switching events: a form of length bias. To assess the effect of task-switching on task completion time it is necessary to estimate task lengths had they not experienced any task-switching, while also accounting for length bias. We review an existing basic semi-parametric homogeneous method and propose modifications incorporating heterogeneity. Each method necessarily uses lengths for tasks unaffected by task-switching to generate estimates. Their performances are compared via a simulation study and the methods are also applied to observational data from a hospital emergency department, where the modified method produces a different test outcome to that of the basic method, indicating the importance of modelling heterogeneity.

Keywords: length bias, clinical work, interruption, task-switching

1. Introduction

Clinical work in the hospital environment involves managing the competing demands of safety, efficiency and quality of care. A characteristic of this work is the frequent interjection of unplanned requests for advice or action, phone calls, pager calls, equipment alarms, etc. - often termed interruptions - that prompt a clinician to change their workflow. If the clinician accepts one of these *prompts* then the unfinished primary task may be suspended in order to switch to the prompting secondary task. The primary task is then resumed at some time after the secondary task has been completed. For example, while prescribing a drug, a doctor is asked advice about a patient by an intern. The doctor pauses the unfinished prescription and discusses the intern's patient for a few minutes before resuming the prescription task. This kind of *task-switching* has been associated with a range of negative effects in experimental studies, including increased risk of error [1] and modified task completion time [1, 2, 3]. Several computer-based experimental studies have assessed resumption lag, that is, the time lag between completion of a secondary task and resumption of the primary task [1, 2, 4]. This article is concerned with assessing the impact of task-switching on primary task length in non-experimental settings. If a task is completed in fragments due to task-switching, the time to complete the primary task may increase, implying a loss of efficiency. Conversely, if primary task length is shortened this may indicate rushing to accommodate the increased demand brought by unplanned secondary tasks, with implications for safety.

[a] *Centre for Health Systems and Safety research, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia*
[b] *Department of Statistics, School of Mathematics and Statistics, University of New South Wales, Sydney, Australia*
[*] *Correspondence to: Scott R. Walter, Level 6, 75 Talavera Road, Macqurie University, NSW 2109, Australia. Email: scott.walter@mq.edu.au*

In assessing the impact of task-switching on primary task lengths, it is not enough to compare lengths of tasks fragmented by task-switching with those for tasks that are not. Prompts resulting in task-switching are more likely to arrive during naturally longer tasks, a manifestation of length bias. This phenomenon has been well documented in the context of drawing samples in which physical length or temporal duration is proportional to the probability of being selected. Cox [5] described this effect in sampling textile fibres to estimate the overall fibre length distribution. Since longer fibres are more likely to be included in a cross sectional sample, the sampled fibre lengths are biased upwards relative to all fibre lengths. If the density of the population of all fibres is f with mean μ then the density of observed fibres, g, is the length biased distribution of f and is given by $g(x) = xf(x)/\mu$. Another well cited example is that of Feller's waiting time paradox [6] where a person intending to catch a bus is more likely to arrive at the bus stop during a longer inter-bus period. Length bias, also referred to as the inspection paradox, has been studied in various other contexts including left truncation when recruiting prevalent subjects for survival analysis [7], genome-wide linkage studies [8], chronic disease screening [9] and estimation of wildlife population density [10].

There is a profusion of studies of prompted task-switching in clinical work, most of which are motivated by the potentially detrimental impact of task-switching on quality of care and patient safety. Of all these studies, the issue of task-switching modifying task completion times is only addressed in one (see [11] and also [12] for the corresponding technical note). We refer to that analysis as the *basic Poisson method*, and we extend this work herein. In the context of sampling, as in the textile example above, an interval can only be selected once, while in clinical work task-switching may occur multiple times during a given primary task. We need to describe the distribution of the length of tasks had they not experienced task-switching, in a way that takes length bias into account. In other words, for a given number of task-switches we need a counterfactual estimate of mean task length assuming no effect. A subsequent comparison with observed task lengths subject to task-switching provides a statistical test of the impact of task-switching on task length. In section 2 we give a more formal definition of the problem, and then in section 3 describe the basic method of [12], which assumes a uniform random onset rate for task-switching. An alternative approach is proposed in section 4 and the two methods are assessed via a simulation study in section 5. An example application is shown in section 6 using data from observations of doctors in an emergency department (ED). Discussion of these methods and their potential for application to real data is provided in section 7.

2. General estimation of the impact of task-switching on task length

Let a *task-switch* denote the point at which a primary task is suspended, in response to some external prompt, to allow completion of a secondary task. *Resumption* is the time at which the primary task is then resumed after completion of the secondary task. We treat the work process as a sequence of time intervals spent on primary tasks, with the possible presence of isolated task-switch points within each primary task interval. Let the i^{th} task length denote either the i^{th} such time interval, or its length T_i. Let the number of task-switches within a task length be the random variable K, and let f_k denote the density function (pdf) of any T_i, conditional upon $K = k$. Because of the length bias effect, the $\{f_k\}$ become stochastically larger as k increases, and our aim is to ascertain whether interruptions make this increase greater, or less than what is expected under length-biasing.

Consider prompted task-switches to be generated by a point process independent of the sequence $\{T_i\}$. Let f be the pdf of task lengths under the null hypothesis, H_0, that task-switches have no effect on total task lengths. Denote the probability that $K = k$ for a given task length $T = t$, under H_0, by $P(k \mid t)$. Then the joint continuous-discrete pdf of T and K is given by $f(t, k) = P(k \mid t)f(t)$, and under H_0,

$$f_k(t) = \frac{f(t,k)}{P(k)} = \frac{P(k \mid t)f(t)}{\int_0^\infty P(k \mid s)f(s)ds} \tag{1}$$

Let f_k^* denote the pdf of T, given $K = k$, when H_0 is not true. Then f_0 and f_0^* will remain equivalent as neither is affected by task-switching, whether or not H_0 is true. But if H_0 is not true then for values $k \geq 1$ we expect that f_k^* is either stochastically larger or stochastically smaller than f_k.

A suitable test statistic for these possible effects, approximately standard normal under H_0, is

$$Z_k = \frac{\hat{\mu}_k - \tilde{\mu}_k}{\sqrt{\tilde{\sigma}_k^2/n_k}},$$

where $\hat{\mu}_k$ is the sample average of the n_k values of T when $K = k$, and where $\tilde{\mu}_k$ and $\sqrt{\tilde{\sigma}_k^2/n_k}$ are its expected value and standard error, respectively, under H_0. In the following sections we show these last two values are expressible in terms of corresponding values for $k = 0$, and can be considered to be known exactly in the example to be discussed in section 6, because of the very large amount of data for $k = 0$. Thus Z_k can be calculated, and under H_0 comes from an approximately $N(0, 1)$ distribution for moderate or large n_k.

3. Previous Methodology - The basic Poisson Method

In considering particular methods that follow the general outline above, we first revisit the approach of [12] which we refer to as the *basic Poisson* method due to its reliance on the assumption that task-switches occur according to a homogeneous Poisson process with constant rate λ per unit of time. The joint distribution of (T, K) under H_0 is then

$$\begin{aligned} f(t,k) &= P(k \mid t)f(t) \\ &= e^{-\lambda t}(\lambda t)^k f(t)/k! \end{aligned}$$

since the distribution of the number of events in a time interval of length t for a Poisson process with rate λ is Poisson with parameter λt.

The density for the time on task given k task-switches is then

$$f_k(t) = \frac{t^k e^{-\lambda t} f(t)}{\int_0^\infty s^k e^{-\lambda s} f(s) ds} \propto t^k e^{-\lambda t} f(t) \tag{2}$$

as a function of t, showing how these densities adjust the null hypothesis task-time density f for the impacts of length biased sampling. Using the notation $\mu_k(r)$ to denote $E(T^r|K = k)$, we have

$$\mu_k(r) = \frac{\int_0^\infty t^{k+r} f_0(t) dt}{\int_0^\infty s^k f_0(s) ds} = \frac{\mu_0(k + r)}{\mu_0(k)} \tag{3}$$

This is the ratio of the $(k + r)^{th}$ to the k^{th} raw moment of T when $k = 0$, and can be approximated by the ratio of the corresponding sample moments for tasks with no task-switching. In the example discussed in section 6 there is a very large amount of data for $k = 0$, so that all moments $\mu_0(r)$ may be considered to be known exactly.

For investigating the effect of task-switching upon task lengths, a convenient test statistic could be any sample moment of T values when $K = k$, for some $k \geq 1$, because its null mean and variance can be calculated from moments when $k = 0$, as just outlined. For example, consider $\hat{\mu}_k$, the sample average of T values when $K = k$, for $k \geq 1$. It has null mean $\mu_k(1) = \mu_0(k + 1)/\mu_0(k)$, and null variance $n_k^{-1}\sigma_k^2$ where there are n_k observations with $K = k$, and

$$\sigma_k^2 = \mu_k(2) - \mu_k(1)^2 = \frac{\mu_0(k + 2)}{\mu_0(k)} - \left\{ \frac{\mu_0(k + 1)}{\mu_0(k)} \right\}^2$$

Note that the value of the Poisson task-switching rate λ does not need to be known.

This method was applied in [12] to observations of doctors in the ED of a large teaching hospital in Sydney, Australia. Expected mean task lengths based on the Poisson assumption were compared with observed mean task lengths for $k = 1, 2, 3+$ and, somewhat counterintuitively, the results suggested that tasks were shortened when affected by task-switching. In that data there was also considerable heterogeneity in the task-switching rate conditional on several variables such as type of task, doctor role and also time of day. The analysis was stratified into three time periods, each of which appeared to have a different rate, and the shortening effect was still observed within each stratum. However, heterogeneity in the rate persisted due to the other non-temporal factors, including task type, indicating that the assumption of a homogeneous Poisson rate was not satisfied, other than within each of the 131 observation sessions. Insufficient sample size within individual sessions limited the possibility of applying the method at session level. Thus, motivated by the need for a more flexible method involving variable Poisson rates, we now examine an alternative and extension to the original basic Poisson method.

4. Alternative approach - The Poisson mixture model

The homogeneous rate assumption of the basic Poisson method is often not met by data from real settings. For the ED data, the variation in the Poisson task-switching rate is related to many contextual factors. There was also variation between observation sessions and it was previously shown that the homogeneous Poisson assumption was reasonable when applied within individual sessions, but not when applied to the aggregated data. The heterogeneity of the rate observed in this data instead suggests Poisson mixtures caused by variable rates. While there are several ways to specify the process governing the rate, we consider the case where the rate parameter is like a random variable from a gamma $\Gamma(\alpha, \beta)$ distribution. In addition to being flexible enough to model many forms of mixing, the gamma distribution is conjugate to the Poisson, allowing algebraic simplifications.

The distribution of Poisson rates mixed according to a gamma distribution gives the following

$$P(k \mid t) = \int_0^\infty \frac{1}{k!}(\lambda t)^k e^{-\lambda t} \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda/\beta} d\lambda = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)k!} \beta^k t^k (1+\beta t)^{-\alpha-k}$$

In this formulation the mixing applies at task level, that is, the i^{th} task with length T_i is assumed to have a distinct rate parameter, $\lambda_i \sim \Gamma(\alpha, \beta)$. This is the rate per unit time and so the number of task-switches for a given task length is then $\text{Pois}(\lambda_i t_i)$, where the overall mean rate is given by $\lambda = \alpha\beta$. This is equivalent to assuming a negative binomial distribution for the number of task-switches per task, with rate λ and dispersion parameter α. Allowing for variation in rate at task-level is likely to be sufficient to capture variation due to the factors observed in the ED study: task type, doctor role and temporal variations. It also provides more flexible assumptions about the task-switching process relative to the Poisson model.

Adapting the derivation of f_k for the basic Poisson model in section 3 to the case of gamma mixing of the rate λ, we obtain

$$f_k(t) = \frac{t^k (1+\beta t)^{-\alpha-k} f(t)}{\int_0^\infty s^k (1+\beta s)^{-\alpha-k} f(s) ds}$$

We can alternatively write this as

$$f_k(t) = \frac{t^k (1+\beta t)^{-k} f_0(t)}{\int_0^\infty s^k (1+\beta s)^{-k} f_0(s) ds}$$

from which a general identity follows

$$E_k\{A(T)\} E_0 \left\{ \frac{T^k}{(1+\beta T)^k} \right\} = E_0 \left\{ \frac{T^k A(T)}{(1+\beta T)^k} \right\} \tag{4}$$

where A denotes any function of task lengths T. Thus E_k quantities are expressed in terms of E_0 quantities. Because of the large amount of data for $k = 0$, E_0 quantities are assumed to be known exactly. Then, making suitable choices of the function A enables the null mean and variance to be calculated for test statistics which are empirical averages of functions of T. A simple choice of A, analogous to the basic Poisson model, is $A(T) = T$. Unlike other choices of A, this provides an estimate of expected task length in addition to a test statistic. $A(T) = T^2$ can be used to generate a variance estimate. Note that for a given λ, $\beta \to 0$ as $\alpha \to \infty$ and this method becomes equivalent to the basic Poisson model.

While the basic Poisson model does not require knowledge of the task-switching rate, the mixture approach must explicitly estimate parameters of the task-switching process. The parameter β may be estimated as $\beta = \lambda/\alpha$ by fitting a negative binomial model to the full collection of tasks to obtain estimates of the overall rate λ and dispersion α. Counts of task-switches per task are modelled with task length as the offset and by fitting an intercept only. Since this procedure uses all tasks, β is considered to be known exactly.

5. Simulation study

We simulated four scenarios that either aligned with or challenged the assumptions of each method. Since neither approach makes assumptions about the distribution of task lengths, tasks in the synthetic data were arbitrarily drawn from an exponential distribution with mean parameter 100. This approximately corresponds to the mean task length, in seconds, of tasks in the example data in section 6. To further approximate the example dataset, we simulated a sample of 10,000 tasks. Task-switches were chosen to occur more frequently in the simulation than in the example to generate sufficient numbers of tasks with three or more task-switches. For the Poisson mixture method we only considered the case where $A(T) = T$. The focus of the simulation study was to assess test validity rather than power, so H_0 was simulated as true.

In all four simulated scenarios, the number of task-switches was treated as originating from a Poisson mixture distribution with the form of the mixing varying between scenarios. For each task, λ_i is drawn from the mixing distribution. Then the number of task-switches is drawn from $\text{Pois}(\lambda_i t_i)$. In the first scenario the mixing distribution is a degenerate distribution with parameter λ having a single value, $1/120$. This is equivalent to assuming task-switching follows a homogeneous Poisson process. In the second scenario, the mixing distribution is $\Gamma(\alpha, \beta)$ consistent with the assumptions of the Poisson mixture model, where $\alpha = 2$ and $\beta = 1/240$. Since for very large α task-switching is approximately homogeneous Poisson, the value of 2 was chosen to be clearly more overdispersed. The third scenario presents a form of mixing simpler than the gamma distribution but more involved than the homogeneous Poisson case. For each task λ_i can be either $1/60$ or $1/240$ with equal probability. The fourth scenario simulates mixing according to a lognormal distribution, as this generates data more dispersed than the either model can accomodate. One thousand replicates were run for each scenario.

Table 1 shows the results of the simulation study. The numbers of simulated tasks without any task-switching, i.e. where $k = 0$, have been included as these indicate the effective sample size of the methods, since both derive E_0 estimates from these tasks. The first scenario with homogeneous Poisson task-switching showed close agreement between simulated mean task lengths and those predicted by each method. For all values of K the null hypothesis is not falsely rejected, although there is some evidence of underestimation that increases with K. In all other scenarios the basic Poisson method shows clear and consistent overestimation, indicating a lack of resilience to any form of mixing. This suggests that tasks with no task-switching, upon which basic Poisson estimates are based, have a heavier tailed distribution when the rate is heterogeneous rather than homogeneous. The moments $\mu(k + r)$ are increasingly inflated as k increases due to the excess of larger values of t, which causes over estimation in the ratios of moments.

The Poisson mixture approach copes well with gamma mixing, as expected, with Z-scores very close to zero for all values of K. It is also more resilient to the simpler form of mixing in scenario 3 compared to the Poisson model, although it consistently overestimates and falsely rejects the null for $K = 3$. When task-switches arrive according to a lognormal mixture of Poisson processes (scenario 4) the Poisson mixture significantly underestimates the simulated mean task lengths.

6. Example study: task-switching in an emergency department

The study for which the Poisson method was originally developed examined the impact of task-switching (termed interruption) on task length in the ED of an Australian hospital [11]. The corresponding technical note discussed the heterogeneity of the task-switching rate between observation sessions. The Poisson assumption was approximately satisfied within each session but there was insufficient data to apply the method to individual sessions.

Table 2 shows the observed mean tasks times along with times predicted by each method. In addition to overall results, the table also shows results stratified by the three time periods identified in the original study as having distinct task-switching rates [12]. The observed mean task lengths are significantly lower than those preditced by the Poisson method at the 0.05 level. The propensity of the basic Poisson model to overestimate predicted values in the presence of mixing, as seen in the simulation study, is consistent with the high predicted values in the example data. In contrast, among results from the Poisson mixture model only one suggested a significant difference. Otherwise, the remaining non-significant effects went in both directions. Where the Poisson model provided strong evidence that tasks were shortened in the presence of tasks-switching, the mixture model does not provide clear evidence of an effect.

7. Discussion

In this study we have extended the methodology on dealing with length bias when assessing the impact of task-switching on task length in clinical work. The main existing method (basic Poisson) has been examined more extensively and the proposed new method (Poisson mixture) offers an alternative that may be more broadly applicable. Both methods are equivalent under the assumption that task-switches arrive according to a Poisson process, but the Poisson mixture model allows more flexibility in the task-switching arrival process providing a clear advantage over the original approach.

Of the experimental studies that estimated resumption lag [1, 2, 4], values ranged from about one to three seconds. An observational study of nurses that used eye-tracking technology recorded resummption lags between 0.1 and 6.7 seconds [15]. If changes in task length of this magnitude occur in clinical work as a result of task-switching then the methods considered herein will be limited in their ability to detect such changes unless the sample size is very large. These studies focused on the cognitive component of resumption lag, that is, the time to taken to collect ones thought in preparation to resume a task. It is possible, however, that in clinical settings the delay in switching back to the original tasks could take longer than a few seconds due to practical considerations. For example, where resumption of the original task requires changing location, logging back into a computer system or locating misplaced equipment. The reinforced significant findings in the example indicate a considerable shortening of tasks after task-switching which may suggest rushing to complete tasks under the time pressure created by intrusion of unplanned prompts [11].

Often in clinical work that involves frequent task-switching, there are many different types of tasks being performed at different times and as many types of prompts that trigger task-switching. Hence a natural focus for further methods is to be able to more felxibly acccomodate different forms. The Poisson mixture extension allows the Poisson rate parameter to vary according to a parametric distribution. Distributions other than Poisson-gamma may be specified, for example the Poisson-lognormal distribution has been discussed elsewhere [13]. The rate parameter can also be a deterministic function of time as per a non-homogeneous Poisson process.

$$P(k \mid t) = \frac{m(t)^k}{k!} e^{-m(t)}$$

where $m(t) = \int_0^t \lambda(u) d\lambda$. The interarrival time distribution for this case is discussed by Yakovlev et al. [19] The rate parameter can also be allowed to vary randomly with time, as is the case for a Cox process. An approach more relvant to observations of clinical work is to model the mixing as a function of covariates that are known to affect the task-switching rate either with a log-linear model, or possibly with generalized linear autoregressive moving average model to incorporate dependence on past events (*William, what's the best ref for this?*). While the explicit modelling of the task-switching rate is appealing, this approach does not have the algebraic simplifications of the Poisson mixture model.

The impact of task-switching on task completion time in clinical work has implications for efficiency and safety, yet only one study attempted to assess this effect. The comparison of the existing method with several new approaches has provided key insights into the complexity of quantifying the task-switching effect while accounting for length bias. The reliance of each method on particular parametric assumptions means that their application is limited to data that satisfies the corresponding assumptions. For the most part these assumptions relate to the task-switching process. A fully non-parametric solution would be more widely applicable to real data, however, the flexible accomodation of mixing in the distribution of task-switching arrivals is a practical semi-parametric alternative.

8. Acknowledgements

This study was funded under the ARC Discovery Grants DP110100090 and DP160100943, and NHMRC Program grant 1054146. SW is funded by a NHMRC Postgraduate Scholarship.

References

1. Bailey BP, Konstan JA. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*; 22: 685-708.
2. Altmann EM, Trafton JG. 2004. Task Interruption: Resumption Lag and the Role of Cues. *Proceedings of the 26th annual conference of the Cognitive Science Society*;57 August 2004, Chicago, IL: Lawrence Earlbaum Associates, Inc., 2004:438.
3. Speier C, Valacich JS, Vessey I. 1999. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*; 30: 337360.
4. Hodgetts HM, Jones DM. 2003. Interruptions in the Tower of London Task: Can Preparation Minimise Disruption? In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 47(8): 1000-1004.
5. Cox DR. 1969 *Some sampling problems in technology*. *In*: New develpments in survey sampling. John Wiley and Son: New York.
6. Feller W. 1971. An Introduction to Probability Theory and its Applications (2nd edn). John Wiley and Son: New York.
7. Huang CY, Qin J. 2011. Nonparametric estimation for length-biased and right-censored data. *Biometrica*; 98(1): 177-186.
8. Terwilliger JD, Shannon WD, Lathrop GM, Nolan JP, Goldin LR, Chase GA, Weeks DE. 1997. True and false positive peaks in genome wide scans: Applications of length-biased sampling to linkage mapping. *American Journal of Human Genentics*; 61: 430-438.
9. Zelen M, Feinlib M. 1969. On the theory of screening for chronic diseases. *Biometrika*; 56(3): 601-614.
10. Patil GP, Rao CR. 1978. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*; 34: 179-189.
11. Westbrook JI, Coiera E, Dunsmuir WTM, Brown BM, Kelk N, Paoloni R, Tran C. 2010. The impact of interruptions on clinical task completion. *Quality and Safety in Health Care*; 19(4):284-289.
12. Brown BM, Dunsmuir WTM. 2010. Adjusting for length biased sampling when assessing the impact of interruptions on total time on task in a hospital emergency department setting. *Quality and Safety in Health Care*; 19(4):284-289. supplement. `http://qualitysafety.bmj.com/content/suppl/2012/05/17/qshc.2009.039255.DC1/Technical_note_revised_Jan_2010.pdf`
13. Bulmer MG. 1974. On Fitting the Poisson Lognormal Distribution to Species-Abundance Data. *Biometrics*; 30(1): 101-110.
14. Chen SX. 2000. Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*; 52(3): 471-480.
15. Grundgeiger T, Sanderson P, MadDougall HG, Venkatesh B. 2010. Interruption management in the intensive care unit: predicting resumption times and assessing distributed support. *Journal of Experimental Psychology: Applied*; 16(4):317-334.
16. Gupta RC. 1979. Waiting time paradox and size biased sampling. *Cummunications in Statistics - Theory and Methods*; 8(6): 601-607.
17. Rivera AJ. 2014. A socio-technical systems approach to studying interruptions: understanding the interrupter's perspective. *Applied Ergonomics*; 45(3): 747-756.
18. Walter SR, Li L, Dunsmuir WTM, Westbrook JI. 2014. Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours. *BMJ Quality and Safety*; 23; 231-241.
19. Yakovlev G, Rundle JB, Shcherbakov R, Turcotte DL. 2008. Inter-arrival time distribution for the non-homogeneous Poisson process. arXiv:cond-mat/0507657v1.

Table 1. Simulated tasks lengths compared to expected lengths generated by the basic Poisson model and Poisson mixture model.

Distribution of no. of task-switches $P(K \mid T)$	No. of task-switches (K)	No. of tasks	Simulated mean task length	Basic Poisson model			Poisson mixture model		
				Expected task length	SE	Z-score	Expected task length	SE	Z-score
1. Exponential	0	5452	54.5	-	-	-	-	-	-
	1	2482	109.0	109.0	1.5	0.02	108.8	1.5	0.1
	2	1127	163.8	163.4	2.8	0.1	162.9	2.8	0.3
	3	513	218.1	217.4	4.7	0.2	216.3	4.7	0.4
	4	232	272.8	269.7	7.4	0.4	268.1	7.4	0.6
	5	106	327.5	317.9	11.0	0.9	315.8	11.0	1.1
2. Poisson-gamma	0	5944	63.8	-	-	-	-	-	-
	1	2146	113.8	135.5	2.2	-9.8	113.8	1.9	0.04
	2	925	156.0	213.4	4.3	-13.2	156.0	3.3	0.01
	3	444	193.4	295.3	7.3	-13.9	193.1	5.3	0.1
	4	229	226.5	377.2	11.0	-13.7	226.7	7.9	-0.02
	5	125	257.1	451.6	14.7	-13.2	257.5	11.2	-0.03
3. Mixture of Pois$(1/60)$ and Pois$(1/240)$	0	5403	59.1	-	-	-	-	-	-
	1	2210	106.1	126.7	2.0	-10.1	106.7	1.8	-0.3
	2	1039	141.8	198.9	3.7	-15.3	146.9	2.9	-1.7
	3	547	171.6	272.2	5.9	-17.0	182.1	4.4	-2.4
	4	312	201.5	343.1	8.3	-17.1	213.5	6.2	-1.9
	5	187	233.3	406.6	10.6	-16.4	242.0	8.4	-1.0
4. Poisson-lognormal	0	4346	53.2	-	-	-	-	-	-
	1	2034	90.4	120.3	2.1	-14.2	79.7	1.5	6.9
	2	1116	117.8	195.8	3.7	-20.9	99.4	2.3	8.1
	3	677	139.0	275.7	5.6	-24.5	115.6	3.1	7.5
	4	442	156.0	353.6	7.3	-27.2	129.8	4.0	6.5
	5	304	170.3	420.6	8.4	-29.9	142.6	5.0	5.5

Note: Task-switch countl distributions are simulated as: 1. Poisson with rate $1/120$; 2. Poisson-gamma with shape $\alpha = 2$ and scale $]beta = 1/240$; 3. equal mixture of two Poisson with means $1/60$ and $1/240$; 4. Poisson-lognormal. Task lengths are simulated as exponential with mean 100 in all cases. Standard errors (SE) are generated from model formulae, not standard deviation of simulated sample.

Table 2. Application of basic Poisson and Poisson mixture models using data from observations of emergency department doctors.

Observation sessions	No. of task-switches	No. of tasks	Observed mean task length	Basic Poisson model			Poisson mixture model		
				Expected task length	SE	Z-score	Expected task length	SE	Z-score
1 to 15	1	145	120.6	269.7	30.1	-5.0	133.6	16.6	-0.8
	2	30	148.2	756.7	95.5	-6.4	242.2	51.6	-1.8
	3	24	298.6	1118	92.9	-8.8	354.4	69.6	-0.8
	4 or more	18	467.8	1303.4	84.8	-9.9	457.6	89.1	0.1
16 to 66	1	263	200.4	299.6	20.5	-4.8	193.8	13.9	0.5
	2	92	326.5	670.1	49.4	-7.0	311.8	29.9	0.5
	3	33	407.5	1005.5	81.8	-7.3	418.9	57.2	-0.2
	4 or more	26	497.8	1225.3	78.1	-9.3	513.0	70.1	-0.2
67 to131	1	204	197.3	352.7	29.4	-5.3	239.8	20.2	-2.1
	2	33	434.5	853.7	117.1	-3.6	417.9	67.9	0.2
	3	7	376.1	1384.0	297.1	-3.4	587.7	171.8	-1.2
	4 or more	6	616.7	1830.5	320.8	-3.8	735.9	202.8	-0.6
All	1	612	180.5	325.3	15.6	-9.3	191.3	9.6	-1.1
	2	155	315.0	782.3	48.7	-9.6	308.4	24.7	0.3
	3	64	363.3	1251.7	88.4	-10.0	417.7	44.6	-1.2
	4 or more	50	501.2	1651.7	104.9	-11.0	515.3	55.4	-0.3

Note: Observed values in rows corresponding to 4 task-switches are an aggregation of tasks with 4 or more task-switches.

Chapter 7

The Poisson mixture model: Application to observations of doctors in multiple hospital settings

7.1 Chapter background

The article in this chapter represents the first full scale application of the Poisson mixture model. It applies the model, as detailed in chapter 6, to several datasets from previous studies of doctors' work to determine the impact of externally prompted task-switching on task completion time. Since the method and its precursor had previously only been applied to one dataset, there was enormous scope to implement it more widely to build understanding of the time modifying effects of interruptive events in clinical work. To that end this study provides new results about the impacts of task-switching in multiple hospital settings. The results address an important question in the study of clinical work, namely the impact of task-switching on task completion time, and also illuminate several ways forward for this direction of inquiry.

This study addresses the fourth objective of the thesis by applying statistical developments to generate new evidence on clinical work in hospitals.

7.2 Assessing the effect of interruptive events on task completion time: a multi-site study

This paper was written for the *International Conference on Healthcare Systems Ergonomics and Patient Safety* (HEPS 2016) following acceptance of a peer-reviewed extended abstract. It has been published in the conference proceedings following the conference itself which took place October 5-7 2016.

Author contributions

SRW conceived of the study, performed the analysis and drafted the paper. JIW and WTMD gave critical input on the study design and on each draft.

Assessing the effect of interruptive events on task completion time: a multi-site study

Walter, Scott R.[1]*, Dunsmuir, William T.M.[2], Westbrook, Johanna I.[1]
[1] *Centre for Health Systems and Safety Research, Australiana Institute of Health Innovation, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia*
[2] *Department of Statistics, School of Mathematics and Statistics, University of New South Wales, Sydney, Australia*
*e-mail: scott.walter@mq.edu.au

Abstract:
Context: The impact of disruptive events on task completion times in clinical work has implications for both safety and efficiency. However, this has received minimal attention to date due to the methodological challenges of evaluating this association in uncontrolled observational settings.
Objectives: The aims of this study were use a newly developed statistical technique, the Poisson mixture model, to assess the impact of interruptions on task completion time, and to determine the extent to which the association differs between settings.
Methodology: Interruptions are conceptualised as prompted task-switches, that is, switching from one task to another in response to some external prompt resulting in the original task being completed in several fragments. The Poisson mixture model was applied to 600 hours of observational data from several hospital settings: emergency departments (ED), intensive care units (ICU) and general wards. The model was used to generate expected mean task lengths assuming no task-switching effect, which were then compared to observed means via a hypothesis test.
Main results: In the ICUs, general wards and one of the two EDs, there was strong evidence that tasks were shorter when completed in fragments due to task-switching. For tasks with one instance of task-switching, decreases were highly significant ($p<0.001$) and ranged from 46 to 152 seconds, or 27% to 42%. For the other of the two EDs there was no evidence of a task-switching effect. Across all settings there was no evidence that tasks increased in length due to task-switching.
Conclusion: This study addresses a persistent gap in knowledge about the impact of interruptive events on clinical work. The shortening of tasks fragmented by task-switching is a significant finding, and provides impetus and direction for further inquiry including the differential impacts of task-switching and the examination of other types of disruptive events.

Keywords: task-switching, interruption, clinical work

1. Introduction

Interruptions have been an ongoing area of research in healthcare, with many studies focusing on their potential to cause error or contribute to inefficiency in clinical practice. However, quantitatively linking interruptions with safety and efficiency outcomes is challenging in the complex non-experimental settings in which clinical work occurs (Walter, Dunsmuir & Westbrook, 2015). Studies to date have reported associations between interruptions and a range of outcomes, including clinician-level effects such as self-reported workload (Weigl, Müller, Vincent, et al., 2012), task-level effects such as failure to resume an interrupted task (Drews, 2007) and the time cost of resuming such tasks (Grundgeiger, Sanderson, MacDougall, et al., 2010), and clinical outcomes such as dispensing errors (Flynn, Barker, Gibson, et al., 1999) or medication administration errors (Westbrook, Woods, Rob, et al., 2010).

One particular outcome that has received relatively little attention is the impact of interruptions on the time taken to complete tasks. Resumption lag – the time taken to reorient back to a task after having been prompted to switch to another task – has been measured in computer-based experiments (Altmann & Trafton, 2004; Monk 2004) and also in an ICU through use of eye tracking software (Grundgeiger, Sanderson, MacDougall, et al., 2010). In applied settings, including healthcare, there is potential for interruptions to reduce task completion time, and also to increase it through mechanisms other than resumption lag. This suggests that the effects of interruptions in uncontrolled settings cannot just be assessed by measuring lag times associated with switching from one task to another.

Identifying changes in task lengths due to interruptive events from direct observation of clinicians is analytically challenging. Longer tasks will naturally have more interruptions even if there is no interruption effect on task length. This form of *length bias* means that comparing tasks with and without task-switching is not appropriate and could generate a false positive result. For tasks fragmented by one or more interruptions we wish to estimate their average length had they not experienced any interruptions, and these expected values need to be specific to the number of interruptions to account for the length biasing effect. Westbrook, Coiera, Dunsmuir, et al. (2010) were the first to tackle this question by developing and applying a statistical method to predict counterfactual task lengths assuming no interruption effect and a constant probability of interruption occurrence over time (see corresponding technical note: Brown & Dunsmuir, 2010). These predicted task lengths can be compared to observed mean task lengths via a hypothesis test to assess the interruption effect. We have developed a new method, the Poisson mixture model, that extends the original approach by allowing more flexible assumptions that better align with the heterogeneous data from direct observation of clinical work (Walter, Brown & Dunsmuir, 2016).

Due to the inconsistent and unclear use of the term ‘interruption’ in both experimental and healthcare literature we use alternative terms to describe the disruptive aspects of clinical work. A *prompt* is an event that has the potential to elicit a change in workflow. In observational studies these are necessarily restricted to observable external events such as phone calls, questions from other staff, and so on. Clinicians switch between tasks prior to task completion for many reasons, one of which is to respond to external prompts. Thus the act of suspending a primary task, having received an external prompt, then addressing the task related to the prompt (secondary task) is considered *externally prompted task-switching*, however, for brevity we use the shorter form *task-switching* in this paper. The purpose of this study was to assess the impact of these events on task completion time.

2. State of the art

The original method of Brown and Dunsmuir, the basic Poisson model, has only been applied in the study for which it was developed. Despite the pervasiveness of various manifestations of length bias in observational data on clinicians’ work, the issue is almost never discussed in the healthcare literature and at times has resulted in studies claiming effects that are potentially just a symptom of this bias (Trbovich, Griffin, White, et al., 2010; Wiegmann, ElBardissi, Dearani, et al., 2007). The key assumption of the original method, that prompted task-switches occur according to a homogeneous Poisson process, is not often satisfied when applied to the typically heterogeneous task-switching rate observed in clinical work. The Poisson mixture model has been developed to specifically capture that heterogeneity in a more appropriate way. It is currently the best available method to assess fundamental questions about the relationship between task-switching and efficiency, and this study represents its first full scale application to observations from healthcare settings.

3. Objectives and Methods

The aims of this study were to apply the Poisson mixture model to more than 600 hours of observational data on doctors in several hospital settings to assess the impact of task-switching on task completion time, and to determine to extent to which that impact differs between settings. The model was applied to data from four studies. Of these, two were conducted in emergency departments (ED), and the first ED study involved 210 hours of observations on 40 doctors (Westbrook, Coiera, Dunsmuir, et al., 2010), while the second followed 36 doctors for 122 hours (Walter, Raban, Douglas, et al., 2016) (ED 1 and ED 2, respectively, in Table 1). Another was conducted in intensive care units (ICU) at two hospitals with 161 hours of observations on 26 doctors (Li, Haines, Hordern, et al., 2015). The fourth study involved 19 doctors observed over 151 hours on four general wards in one hospital (Westbrook, Ampt, Kearney, et al., 2008).

Each study used the same observational methodology, namely a workflow time study approach (Lopetegui, Yen, Lai, et al., 2014) implemented with the Work Observation Method By Activity Timing (WOMBAT) software on a handheld tablet (Westbrook & Ampt, 2009). Doctors were observed directly and time-stamped information about their activities was recorded according to predefined task categories.

Table 1. Summary of data sources.

Study	Departments	Hospitals	Participants	Hours observed	Task-switching rate (per hour)
ED 1	1	1	40	210	6.0
ED 2	1	1	36	122	5.4
ICU	2	2	26	161	3.5
Wards	4	1	19	151	2.9

The rate of prompted task-switching is known to vary considerably with factors such as task-type or between individual doctors (Walter, Li, Dunsmuir, et al., 2014). In the Poisson mixture model each task is assumed to have a different task-switching rate per unit time, and those task-specific rates are assumed to follow a gamma distribution to capture the heterogeneity of the task-switching rate. Hence the task-switching process is conceptualised as a mixture of rates at task level, rather than a homogeneous rate as per the original approach of Brown & Dunsmuir (2010). The statistical details of the model are described by Walter, Brown & Dunsmuir (2016), but we give a brief outline here.

The Poisson mixture model estimates the expected mean task lengths, $\tilde{\mu}_k$, for a given number of task-switches, k, under the assumption that task-switching has no effect on task length. The estimator of the mean for each value of k is based on a ratio of sample moments, and the expected variance of task lengths, $\tilde{\sigma}_k^2$, is calculated in a similar way,

$$\tilde{\mu}_k = \frac{\sum_{i=1}^{n_0} t_{0i}^{k+1}(1+\beta t_{0i})^{-k}}{\sum_{i=1}^{n_0} t_{0i}^{k}(1+\beta t_{0i})^{-k}}\ ;\ \tilde{\sigma}_k^2 = \frac{\sum_{i=1}^{n_0} t_{0i}^{k+2}(1+\beta t_{0i})^{-k}}{\sum_{i=1}^{n_0} t_{0i}^{k}(1+\beta t_{0i})^{-k}} - \tilde{\mu}_k^2.$$

From this it can be seen that estimates of the expected lengths of tasks with one or more task-switches are derived from lengths of tasks unaffected by tasks switching, represented by t_{0i}, of which there are n_0 such tasks. In addition to being impervious to any task-switching effect, these tasks conveniently also tend to be the most numerous. To obtain a value for the β term, an intercept only negative binomial model is fitted to the task level data and the estimated overall rate $\hat{\lambda}$ and dispersion $\hat{\alpha}$ are used to estimate β using $\hat{\beta} = \hat{\lambda}/\hat{\alpha}$. These expected mean and variance estimates then enable comparison of observed mean task lengths, $\hat{\mu}_k$, and expected mean task lengths via a standard Z test (assuming asymptotic normality for sufficiently large samples):

$$Z = \frac{\hat{\mu}_k - \tilde{\mu}_k}{\sqrt{\tilde{\sigma}_k^2/n_k}}$$

where the denominator is the standard error (SE) of expected task lengths. When calculating observed task means, the length of each task completed in fragments is considered to be the sum of those time fragments. A significant p-value provides evidence that task-switching has an effect on the time taken to complete tasks.

Since the four datasets also contained information on multitasking, adjustment was made to avoid multiple counting of time due to overlapping tasks and to avoid multiple counting of task-switches where they occurred during two or more tasks progressing simultaneously. Although some individual tasks had more than 10 instances of task-switching these were very rare and analyses were restricted to tasks with up to three task-switches.

4. Results

There was no evidence of a task-switching effect for the first ED study. Specifically, there was no difference in observed and expected task lengths for tasks with one switch, while for tasks with two or three switches differences went in opposite directions and neither was significant (Table 2). In contrast, observed task lengths in the second ED study were significantly shorter than expected for tasks with either one or two task-switches. In the ICU data, observed tasks were shorter than expected for all numbers of task-switches considered, but this was only significant where there was one task-switch despite large absolute and relative differences for two or three switches. Similarly for doctors on general wards, observed task lengths were shorter than expected but this was only significant for tasks with one task-switch.

Table 2. Application of the Poisson-mixture model to assess the effect of task-switching on task completion time.

Data source	Number of task-switches	Number of tasks	Mean task length			Comparison	
			Observed	*Expected (SE)*	*% Difference*	*Z score*	*P-value*
ED 1	1	519	193.7	193.7 (10.6)	0.0	-0.003	0.997
	2	153	315.8	308.8 (25.2)	2.2	0.3	0.78
	3	62	369.3	418.4 (45.9)	-11.7	-1.1	0.29
ED 2	1	521	116.9	162.9 (10.9)	-28.2	-4.2	<0.001
	2	72	198.2	321.8 (41.7)	-38.4	-3.0	0.003
	3	18	424.4	465.2 (98.9)	-8.8	-0.4	0.68
ICU	1	375	133.3	181.6 (16.7)	-26.6	-2.9	0.004
	2	26	233.5	408.0 (107.6)	-42.8	-1.6	0.11
	3	10	337.0	672.3 (230.6)	-49.9	-1.5	0.15
Wards	1	274	208.6	360.5 (42.2)	-42.1	-3.6	<0.001
	2	21	448.0	805.3 (226.6)	-44.3	-1.6	0.11

5. Discussion

This study represents the first application of a unique statistical method that assesses a fundamental question in the observational study of clinical work. These results indicate a shortening of tasks in the presence of task-switching in three out of the four settings, while there was no significant evidence that tasks increased in length due to task-switching in any setting. The statistically significant decreases ranged in magnitude from 46 seconds to two and a half minutes and relative decreases from 26.6% to 42.1%. These effects represent considerable changes within a clinical work environment.

The results suggest no immediate loss of efficiency due to task-switching. Rather the findings suggest the opposite, since time costs related to task-switching appear overshadowed by some other mechanism that decreases task length. The authors of the first ED study suggested that a shortening effect may indicate that doctors compensate for a perceived loss of time due to task-switching (Westbrook, Coiera, Dunsmuir, et al., 2010) and this has also been indicated in experimental studies. Monk (2004) found that resumption times were faster on average with increasing numbers of interruptions, and frequent interruptions did not result in increased task length. Also, several studies found a decrease in the lengths of interrupted tasks with no loss of quality (Mark, Gudith & Kocke, 2008; Zijlstra, Roe, Leonova, et al., 2010), while others found a similar decrease only for simple tasks but not for complex tasks (Burmistrov & Leonova, 2003; Speier, Valacich & Vessey, 1999).

Rushing or corner cutting to compensate for having to switch tasks to deal with prompts would clearly be a safety issue as it could increase the risk of various forms of error. However, the experimental evidence discussed above suggests that task shortening occurs in scenarios of relatively low cognitive load without loss of quality, while task lengths do not shorten under high cognitive demand. In other words, there is more flexibility to vary the pace of work when there is spare cognitive capacity. A similar phenomenon has been observed in other occupational settings where the work rate varied in response to time constraints (Latham & Locke, 1975).

The results of this study may therefore represent doctors adjusting to evolving workload demands, in the form of prompts, under largely manageable cognitive load conditions. Healthcare professionals are known to use strategies to manage work demands through interleaving, prioritising and sequencing of tasks (Grundgeiger & Sanderson, 2009). They are likely habituated to the typical prompt types and have honed strategies over time to deal with them, thus mitigating their cognitive impact. The lack of evidence of a task-switching effect for the first ED study may be related to workload. Although workload measures were not available across settings for comparison, the first ED study had the highest interruption rate, a measure associated with workload (Weigl, Müller, Vincent, et al., 2012).

The experiments cited above reported an increased emotional cost, despite participants maintaining their quality while working quicker. This included increases in perceived effort (Zijlstra, Roe, Leonova, et al., 2010), stress, frustration and time pressure (Mark, Gudith & Klocke, 2008). So although the observed task shortening may indicate resilient and flexible clinical practice, task-switching could still contribute indirectly to error risk through its influence on affective state.

When interpreting these results it is important to consider the way in which task intervals are delineated during the observation process (Walter, Dunsmuir & Westbrook, 2015). There may be differences in the way observers and doctors perceive the start, end and switch points of tasks. Therefore the mechanics of data collection may influence estimated task-switching effects and this needs to be taken into consideration in the design of future studies.

The combination of the Poisson mixture model and Z test, has some strengths and limitations. The power to detect task-switching effects is limited by the sample, n_k, of tasks for each number of task-switches. As seen in Table 2, tasks with one switch are the most numerous and this rapidly decreases for tasks with two or three switches. A small number of tasks results in a smaller Z score, and hence reduced significance of the comparison test, despite the absolute difference often being considerable. On the plus side, the method has the potential to be applied to events other than externally prompted task-switching. It could easily be applied to other forms of task-switching since clinicians frequently switch between tasks of their own volition, that is, internally prompt task-switching. Also it could be applied to events that don't necessarily cause a task to be completed in fragments, for example, it could assess the effect of multitasking on task completion time.

Another important result of this study is that it highlights several directions for further research. Given the potential for task-switching to have variable effects on different task types or on different individuals, analyses stratified by such factors would provide more nuanced detail about the impact of task-switching. The experimental evidence indicating that cognitive load can modify the effect of task-switching on task length suggests a need for explicit workload assessment to determine the influence of high and low workload on task-switching effects.

6. Conclusion & perspectives

This study addresses a persistent gap in knowledge about the impact of interruptive events on clinical work. While the results give a strong and consistent indication of the effects of prompted task-switching on task length, they do not give a comprehensive picture of their safety and efficiency implications. It may be that we have evidence of resilient practice where clinicians adapt to constantly evolving demands, rather than cutting corners or rushing. That tasks appear to get shorter when they are fragmented by task-switching across multiple settings is a significant result, but this opens up many questions, providing impetus and direction for further inquiry. In particular, how task-switching has differential impacts across task types, individuals and workload intensities, and the role of the observation process in influencing estimated results.

Acknowledgements

This work was supported by NHMRC grants 1055498 and 1054146, and ARC grant DP160100943.

References

Altmann, E. M., & Trafton, J. G. (2004). Task interruption: Resumption lag and the role of cues. *Proceedings of the 26th annual conference of the Cognitive Science Society*, Chicago, IL (p.438). Lawrence Earlbaum Associates, Inc.

Brown, B.M. & Dunsmuir, W.T.M. (2010). Adjusting for length biased sampling when assessing the impact of interruptions on total time on task in a hospital emergency department setting. *Quality and Safety in Health Care*,

19(4), 284-289. Web only data.
Burmistrov, I., & Leonova, A. (2003). Do interrupted users work faster or slower? The microanalysis of computerized text editing task. *Human-Computer Interaction: Theory and Practice (Part I)–Proceedings of HCI International*, 1, 621-625.
Drews, F.A. (2007). The frequency and impact of task interruptions in the ICU. *Proceedings of the human factors and ergonomics society annual meeting*, 51, 683-686.
Flynn, E.A., Barker, K.N., Gibson, J.T., Pearson, R.E. & Berger, E.A. (1999). Impact of interruptions and distractions on dispensing errors in an ambulatory care pharmacy. *American Journal of Health-System Pharmacy*, 56, 1319-1325.
Grundgeiger, T. & Sanderson, P. (2009). Interruptions in healthcare: Theoretical views. *International Journal of Medical Informatics*, 78, 293-307.
Grundgeiger, T., Sanderson, P., MacDougall, H.G. & Venkatesh, B. (2010). Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *Journal of Experimental Psychology: Applied*, 16(4), 317-334.
Latham, G. P., & Locke, E. A. (1975). Increasing productivity and decreasing time limits: A field replication of Parkinson's law. *Journal of Applied Psychology*, 60(4), 524-526.
Li, L., Hains, I., Hordern, T., Milliss, D., Raper, R., & Westbrook, J. (2015). What do ICU doctors do?: A multisite time and motion study of the clinical work patterns of registrars. *Critical Care and Resuscitation*, 17(3), 159-166.
Lopetegui, M., Yen, P. Y., Lai, A., Jeffries, J., Embi, P., & Payne, P. (2014). Time motion studies in healthcare: What are we talking about? *Journal of biomedical informatics*, 49, 292-299.
Mark, G., Gudith, D., & Klocke, U. (2008). The cost of interrupted work: more speed and stress. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 107-110). ACM.
Monk, C.A. (2004). The effect of frequent versus infrequent interruptions on primary task resumption. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, 295-299.
Speier, C., Valacich, J.S., & Vessey, I. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, 30(2), 337-360.
Trafton, J.G & Monk, C.A. (2007). Task interruptions. *Reviews of Human Factors and Ergonomics*, 3(1), 111-126.
Trbovich, P.L., Griffin, M.C., White, R.E., Bourrier, V., Dhaliwal, D. & East, A.C. (2013). The effects of interruptions on oncologists' patient assessment and medication ordering practices. *Journal of Healthcare Engineering*, 4(1), 127-144.
Walter S.R., Li, L., Dunsmuir, W.T.M. & Westbrook, J.I. (2014). Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours. *BMJ Quality and Safety*, 23, 231-241.
Walter, S.R., Raban, M.Z., Dunsmuir, W.T.M., Douglas, H.E. & Westbrook, J.I. (2016). Emergency doctors' strategies to manage competing workload demands in an interruptive environment: an observational workflow time study. *Manuscript resubmitted for publication*.
Walter, S.R., Dunsmuir, W.T.M., & Westbrook, J.I. (2015). Studying interruptions and multitasking in situ: the untapped potential of quantitative observational studies. *International Journal of Human-Computer Studies*, 79, 118-125.
Walter, S.R., Dunsmuir, W.T.M. & Brown, B.M. (2016). Assessing the impact of task-switching on completion of clinical tasks in the presence of length bias. *Manuscript submitted for publication*.
Weigl, M., Müller, A., Vincent, C., Angerer, P. & Sevdalis, N. (2012). The association of workflow interruptions and hospital doctors' workload: a prospective observational study. *BMJ Quality and Safety*, 21, 399-407.
Wiegmann, D.A., ElBardissi, A.W., Dearani, J.A., Daly, R.C. & Thoralf, M.S. (2007). Disruptions in surgical flow and their relationship to surgical errors: An exploratory investigation. *Surgery*, 142, 658-665.
Westbrook, J. I., Ampt, A., Kearney, L., & Rob, M. I. (2008). All in a day's work: an observational study to quantify how and with whom doctors on hospital wards spend their time. *Medical Journal of Australia*, 188(9), 506-509.
Westbrook JI & Ampt A. (2009). Design, application and testing of the Work Observation Method by Activity Timing (WOMBAT) to measure clinicians' patterns of work and communication. *International Journal of Medical Informatics*, 78, S25–33.
Westbrook, J.I., Coiera, E., Dunsmuir, W.T., Brown, B.M., Kelk, N., Paoloni, R., & Tran, C. (2010). The impact of interruptions on clinical task completion. *Quality and Safety in Health Care*, 19(4), 284-289.
Westbrook, J.I., Woods, A., Rob, M.I., Dunsmuir, W.T.M. & Day, R.O. (2010). Association of interruptions with an increased risk and severity of medication administration errors. *Archives of Internal Medicine*, 170, 683-690.
Zijlstra, F.R.H., Roe, R.A., Leonova, A.B. & Kretiet, I. (2010). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, 72, 163–185.

Chapter 8

Discussion

8.1 Summary

The work of clinicians in hospitals is driven by forces at many levels. It is often fragmented, can evolve in unpredictable ways, and is performed within time and resource constraints. At face value this throws up questions about the quality of care when doctors and nurses are working under cognitively taxing, safety critical conditions. The disruptive and fragmented nature of clinical practice intuitively suggests a need to focus on the potential for negative outcomes. However, the complexity of such work, as yet not fully fathomed, demands an appropriately nuanced approach to its study, in terms of how we conceptualise that work and how we observe it and analyse it. This thesis set out to promote improved understanding of the patient safety implications of the clinical work process. Through augmenting the conceptual nuance underpinning the quantitative study of clinical work, and enhancing the appropriateness and sophistication of observational and statistical approaches, this research has generated results with greater granularity and clinical-relevance, and has examined important aspects of clinical work not previously quantified. In addition to the contribution of these original findings, this work also provides a foundation for increased rigour in the quantitative observational study of clinical work. In this chapter the key contributions of this research are discussed in reference to the thesis objectives followed by some final concluding comments.

8.2 Original contributions

The critical review in chapter 2 of observational and statistical methods used in the study of clinical work was a natural start point for the thesis and addressed the first thesis objective. Moreover, it was an important discourse to contribute to the literature. Several authors have provided editorial discussion of some of the methodological challenges facing the study of clinical work, including issues relevant to the predominant quantitative observational approaches (Coiera, 2012; Westbrook, 2014; Grundgeiger et al., 2016). However, there is little detailed theoretical or practical discussion of these challenges. The known issues with observational studies in general (Sauerbrei et al., 2014) are all the more relevant in this field given the complexity of clinicians' work in hospital settings. The article in chapter 2 particularly highlights the many statistical issues that have a significant bearing on the validity of study results, but are often not addressed. It also offers some of the first detailed practical discussion of ways to address these issues in order to bring more rigour to the way such studies are designed, carried out and analysed. The paper has influenced further discussion of methodological issues by other authors (Grundgeiger et al., 2016), and also forms a basis for ongoing work by the candidate and collaborators in developing theory and practical techniques to address issues outlined in, but not covered by, this thesis.

Discussion of the way aspects of clinical work are defined has been ongoing for two decades. Proposed ways to overcome issues of inconsistency and imprecision in definitions has tended to focus on moving towards consensus on universal definitions, particularly a universal definition of an interruption (McFarlane, 1997; Brixey et al., 2007; Sasangohar et al., 2012). In addressing the second thesis objective, a fundamental contribution of this work was the proposal of an alternative idea: that universal definitions are neither possible nor of practical use, but should be context-specific and as precise as possible. Based on this idea, the development and application of a context appropriate system of terms and definitions in this thesis represents a significant advance on previous conceptualisations of the clinical work process. The construction of clinical work as a series of states and transitions between those states forms building blocks from which clinically relevant definitions could be developed, such as the prompt-response process applied in chapters 4 and 5. It allows many existing definitions from both experimental studies and healthcare contexts to be mapped onto specific sequences of states and transitions. It also supports a whole suite of analytic techniques applicable to series of states that have not been applied to clinical work, including Markov models, Bayesian networks and queueing theory.

In many fields of inquiry involving complex human systems, proportionally complex quantitative techniques are used. For example, the proliferation of machine learning techniques being applied to medical diagnosis or online consumer behaviour. In contrast, the methods used in the quantitative study of clinical work have largely been at the simple end of the analysis spectrum relative to the complexity of the setting. In line with the third objective, this thesis has proposed new ways of analysing data from observations of clinical work and has implemented existing techniques in novel ways in order to progress the analytic standards and possibilities in the field. Fundamental ideas about validity were reinterpreted in the context of studying clinical work through workflow time studies, including the need to capture and account for the array of potential confounders at multiple levels of the work system, and to address multiple forms of correlation within outcome measures. The thesis also contains one of the first, if not *the* first, applications of nonparametric regression to the analysis of clinical work. This highlights the facility of such techniques for exploratory analysis in generating a more granular understanding of clinical work from which targeted hypotheses can be derived.

The developments in the statistical analysis of workflow time study data combined with the new conceptualisation of clinicians' work has generated important new evidence about everyday clinical work practices in line with the fourth thesis objective. In particular, results from the applied studies in chapters 4, 5 and 7 are unique in terms of the number of factors examined at multiple system levels, the comparison of effects between hospital settings, and the fact that they address important, but barely explored, research questions. The papers in chapters 4 and 5 both identified a diverse set of factors that had a significant influence on clinicians' strategies to handle external prompts. These included aspects of individual tasks, characteristics of observed clinicians and those they interact with, temporal factors, preceding strategy choices, and the clinician's physical location within the work environment. Such quantitative identification of factors at multiple levels had not previously been done. The latter study also provided the first detailed characterisation of a broad range of clinicians' strategies - task-switching, multitasking, acknowledgement, etc. - as made possible by the development of the prompt-response concept. Also in that study, prompts were observed to be a central conduit for advice provision and information transfer. The study in chapter 7 is one of the very few to assess the effect of task-switching on task completion time and is the first to do so across multiple hospital settings. The observed reduction in average task length in the presence of task-switching was interpreted in light of experimental evidence as indicating that clinicians adjust their rate of work in response to evolving workload demands.

Two main themes have emerged from the findings of the three applied studies. First, the results underscore the point made in the introductory chapter: that clinical work in hospitals occurs in complex settings, where there are many interrelated factors influencing clinicians' work at multiple levels. Secondly, there is emerging evidence against the predominant perception that the disruptive aspects of clinical work are negative. Rather, to some extent they may perform important functions in the delivery of quality care under time and resource constraints, and clinicians may be resilient to their potential to contribute to error or inefficiency.

8.3 Conclusions

The study of clinical work processes is a unique area of research that lies at the intersection of many disciplines. Advancing our understanding of the patient safety implications of the disruptive aspects of these processes requires generating evidence via multiple methods of inquiry, both qualitative and quantitative. The quantitative observational approach, particularly the workflow time study, has vast potential to contribute important knowledge, but has been underexplored due to the methodological challenges of studying complex uncontrolled settings. This thesis has made significant progress in addressing those challenges. Through the application of improved observational and statistical methods it has progressed debates about the conceptualisation of clinical workflow, identified factors that influence clinicians' strategies to manage disruptive events in a range of healthcare settings, and better quantified the impact of interruptions on task completion time. This has provided a more sophisticated understanding of the relationships between work behaviours, work efficiency and error production. Moreover, the methodological progress enables future creation of knowledge necessary for safety improvement.

References

Brixey JJ, Robinson DJ, Johnson CW, Johnson TR, Turley JP, Zhang J. 2007. A concept analysis of the phenomenon interruption. *Adv Nurs Sci*, 30(1): E26-E43.

Coiera E. 2012. The science of interruption. *BMJ Qual Saf*, 21(5): 357-360.

Grundgeiger T, Dekker S, Sanderson P, Brecknell B, Liu D, Aitken LM. 2016. Obstacles to research on the effects of interruptions in healthcare. *BMJ Qual Saf*, 25: 392-395.

McFarlane DC. 1997. Interruption of people in human-computer interaction: a general unifying definition of human interruption and taxonomy. Naval Research Laboratory: Washington, DC, USA.

Sasangohar F, Donmez B, Trbovich P, Easty AC. 2012. Not all interruptions are created equal: positive interruptions in healthcare. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2012, 56: 824-828.

Sauerbrei W, Abrahamowicz M, Altman DG, le Cessie S, Carpenter J. 2014. STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative. *Stat Med*, 33: 5413-5432.

Westbrook JI. 2014. Interruptions and multi-tasking: moving the research agenda in new directions. *BMJ Qual Saf*, 23: 877-879.

Appendix A

Study documents

A.1 Ethics approval letters

The data used in the analysis in chapter 5 comes from a broader study of emergency department doctors conducted in a large tertiary hospital in Sydney, Australia. Ethics approval for the study was first sought from the relevant human research ethics committee overseeing the hospitals in the area, and final approval was granted in May 2014. Approval was then sought from the research governance office of the hospital itself which was subsequently granted in June 2014, after which time observations were able to commence. Copies of both approval letters are included in this section.

Health
South Eastern Sydney
Local Health District

HUMAN RESEARCH ETHICS COMMITTEE
Room G71 East Wing
Edmund Blacket Building
Prince of Wales Hospital
RANDWICK NSW 2031
Tel: 02 9382 3587 Fax: 02 9382 2813
RSOSESLHD@SESIAHS.HEALTH.NSW.GOV.AU
www.seslhd.health.nsw.gov.au/POWH/researchsupport

21 May 2014

Mr Scott Walter
Australian Institute of Health Innovation
Level 1 AGSM Building (G27)
University of New South Wales
UNSW SYDNEY NSW 2052

Dear Mr Walter

HREC ref no: 13/310 (HREC/13/POWH/674)
Project title: Clinicians' Strategies to Manage Work in Emergency Care Settings

Thank you for your correspondence dated 9 May 2014 to the Human Research Ethics Committee (HREC) responding to questions which arose at the Executive Committee meeting on 22 April 2014.

The Executive Committee reviewed your response on 20 May 2014.

Ethical approval has been given for the following:

- Response Letter, dated 9 May 2014
- Amendment Form, dated 9 May 2014
- Protocol, version 3.1, dated 9 May 2014
- Participant Information Sheet, version 3.1, dated 9 May 2014
- Additional Information Email, dated 5 May 2014

Ethical approval is valid for the following site(s):

- Prince of Wales Hospital

This amendment has also been reviewed by the Research Governance Officer at SESLHD. Further authorisation of the above approved documents is not required

for any site that has the Research Governance conducted by the SESLHD Research Support Office. Implementation of this amendment can now proceed.

For multi-site projects reviewed by the HREC after 1 January 2011 a copy of this letter must be forwarded to all Principal Investigators at every site approved by the SESLHD HREC for submission to the relevant Research Governance Officer along with a copy of the approved documents.

Should you have any queries, please contact the Research Support Office on (02) 9382 3587. The HREC Terms of Reference, Standard Operating Procedures, membership and standard forms are available from the Research Support Office website: http://www.seslhd.health.nsw.gov.au/POWH/researchsupport/default.asp

Please quote **HREC ref no 13/310** in all correspondence.

We wish you every success in your research.

Yours sincerely

Deborah Adrian

Deborah Adrian
Executive Officer, Human Research Ethics Committee

This HREC is constituted and operates in accordance with the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research (2007)*, NHMRC and Universities Australia *Australian Code for the Responsible Conduct of Research (2007)* and the *CPMP/ICH Note for Guidance on Good Clinical Practice.*

Health
South Eastern Sydney
Local Health District

RESEARCH SUPPORT OFFICE

Room G71, East Wing
Edmund Blacket Bldg
Prince of Wales Hospital
Cnr High & Avoca Streets
RANDWICK NSW 2031
Tel: (02) 9382 3587
Fax: (02) 9382 2813

25 June 2014

Mr Scott Walter
Australian Institute of Health Innovation
Level 1 AGSM Building
University of NSW, NSW 2052

Dear Mr Walter

RE: SSA Ref: **14/G/129**
HREC / AURED Ref: **13/310 HREC/13/POWH/674**
Project Title: **Clinicians' Strategies to Manage Work in Emergency Care Settings.**

I refer to your Site Specific Assessment application for the above titled project. I am pleased to advise that on 23 June 2014, the Director of Operations granted authorisation for the above project to commence at the Prince of Wales Hospital.

The following conditions apply to this research project. These are additional to any conditions imposed by the Human Research Ethics Committee that granted ethical approval:

1. Proposed amendments to the research protocol or conduct of the research which may affect the ethical acceptability of the project, and are submitted to the lead HREC for review, are copied to the Research Governance Officer.

2. Proposed amendments to the research protocol or conduct of the research which may affect the ongoing site acceptability of the project are to be submitted to the Research Governance Officer.

If you have any queries relating to the above please contact the Research Support Office on 9382 3587.

Yours sincerely

Dr Tali Leizer
Research Governance Officer

SSA 14-129 - Mr Scott Walter - Approval Ltr 25-6-2014.doc

A.2 Study protocol

As part of the ethics approval process, it was necessary to prepare a detailed study protocol describing objectives, all aspects of the design, the proposed statistical analysis as well as safety considerations including confidentiality. The document included in this section of the appendix is the final approved protocol.

Clinicians' Strategies to Manage Work in Emergency Care Settings

Study Protocol

Version 3.1
9th May 2014
HREC Ref: 13/310

Principal Investigator: ***Scott Walter***
Centre for Health Systems and Safety Research, University of New South Wales
Level 1 AGSM Building
University of New South Wales
Kensington NSW 2052
Ph: (02) 9385 0510
Fax: (02) 9385 8280
scott.walter@unsw.edu.au

Co-Investigator: ***Professor Johanna Westbrook***
Centre for Health Systems and Safety Research, University of New South Wales
Level 1 AGSM Building
University of New South Wales
Kensington NSW 2052
Ph: (02) 9385 3856
Fax: (02) 9385 8280

Co-Investigator: ***Professor William Dunsmuir***
School of Mathematics and Statistics, University of New South Wales
Level 2 Red Centre
University of New South Wales
Kensington NSW 2052
Ph: (02) 9385 7035
Fax: (02) 9385 7123

Co-Investigator: ***Magda Raban***
Centre for Health Systems and Safety Research, University of New South Wales
Le
vel 1 AGSM Building
University of New South Wales
Kensington NSW 2052
Ph: (02) 9385 3918
Fax: (02) 9385 8280

Co-Investigator: ***Dr Heather Douglas***
Centre for Health Systems and Safety Research, University of New South Wales
Level 1 AGSM Building
University of New South Wales
Kensington NSW 2052
Ph: (02) 9385 3592
Fax: (02) 9385 8280

Co-Investigator: ***Dana Strumpman***
Pharmacy Department, Prince of Wales Hospital
Barker Street
Randwick NSW 2031

Summary

Title	Clinicians' strategies to manage work in emergency care settings
Protocol version	Version 3 – 11th April 2014
Primary objectives	**1.** To determine how the type of prompt affects which work management strategy a clinician uses in response. **2.** To assess the relationship between workload and the use of work management strategies. **3.** To examine the aspects of work that may contribute to the occurrence of memory lapses. **4.** To compare the rate of prescribing errors in the presence vs. absence of external prompts. **5.** To assess the influence of Working Memory Capacity and polychronicity on the selection of work management strategies. **6.** To determine the effect of workload in modifying the relationship between Working Memory Capacity and work management strategies
Secondary objective	**1.** To use the collected data to develop new quantitative approaches for analysing observational time and motion data.
Study design	Single site observational time and motion study of emergency department doctors
Duration of the study	Data collection: May-Oct 2014 Analysis and write up: Oct 2014 – Dec 2015

TABLE OF CONTENTS

1. Background

1.1 General background

Hospital work involves many competing demands that clinicians need to manage, including safety, efficiency and quality of care. This is even more apparent in critical care settings such as the emergency department. This balancing act requires constant decisions about how best to manage work, a part of which involves dealing with interruptions and performing tasks concurrently, or multitasking.

Research from the field of experimental psychology has shown that interruptions can have negative effects, including increased error, stress, and task completion time (1-3). However, despite the many studies on doctors' and nurses' work, our understanding of interruptions and multitasking in clinical work is far from complete (4). Studying these aspects of work is almost prohibitively complex and the majority of studies to date have been largely descriptive or have examined relatively simple associations.

Some recent studies have begun to offer an expanded view of interruptions and multitasking that focuses on particular actions taken by clinicians in response to *prompts* in the environment, for example questions from other staff, phone calls, pager calls, etc. Based on possible response strategies proposed by others (5-8), we define four observable work strategies which clinicians use in response to prompts, namely *task-switching*, *multitasking*, task *deferral* and *deflection*. Under this framework, an interruption as an externally driven event is reconceptualised as task-switching, a strategy determined by the clinician in response to a prompt. In this case the prompt triggers the clinician to switch from one task to another. Multitasking, that is, concurrent task performance, may be triggered by a prompt or may be initiated by the clinician. This is a lesser studied phenomenon in clinical settings and those studies that exist are mostly descriptive. Deferral is the postponing of action while deflection is the repulsion of a task, both of which are typically achieved by a short communication: “I'll do it later” or “ask doctor X to do it”. Neither of these strategies have been studied in detail. Understanding the full spectrum of strategies that clinicians use to juggle competing demands is important for the development of improvements to clinical work management practices, including error mitigation.

1.2 Rationale for performing the study

This study builds on previous work by the principal and co-investigators (9-12) to examine all observable work management strategies, not just task-switching and multitasking, in order to gain a more comprehensive understanding of the flow of work, the ways in which clinicians manage it, and the potential for adverse outcomes to occur as a result. The observational time-and-motion approach used in these previous studies will be applied to the proposed study to collect new data and answer distinct research questions.

A recent study by the investigators used the idea that upon being prompted, clinicians have a range of response options (9). Using the same idea, but also capturing all observable strategies as well as more detail on prompts, we can gain a more complete picture of the determinants of what strategy a clinician uses in a given situation. If particular strategies are associated with potential for error then identifying factors that initiate those strategies can inform the design of safer work practices.

Error mitigation is often the motivating focus in studies of clinicians' work strategies. Several studies have noted ecological associations between task-switching rates and error rates, but determining association between particular tasks (or sequences of events) and particular errors is much more complicated, only one study has attempted to do so in a clinical setting

(11). A form of memory lapse that is observable within the proposed approach of this study is failure to follow up on deferred action. A clinician may deal with a prompt by deferring some kind of action until a later time, however, due to events in the intervening period the clinician may forget to follow up on the promised action. A diary study of German nurses reported an association between interruptions and forgetting of intentions (13), however, this association has not been examined among doctors nor in critical care settings. We also intend to examine failures to resume suspended tasks as another form of memory lapse.

Prescribing errors are another observable outcome and are a relatively frequent occurrence in hospitals, including the ED.(14-18) Such errors have the potential to negatively impact patient safety (17, 19-21). The investigators have previously documented prescribing errors on wards across a number of hospitals in Sydney (17, 22). Since the risk of prescribing error may be increased by the occurrence of prompts during a prescribing task(15, 16, 23), this study will build on previous work to examine this relationship. Several ED performance indicators measure the provision of timely treatment to patients. Time to analgesia is one such indicator, and may be negatively impacted by the frequency and nature of external prompts. This relationship will be examined alongside the assessment of prescribing errors.

In previous work, a high level of variation between the way individuals handle prompts has been observed (9), and some of this variation is likely due to individual differences. Working memory is broadly defined as a general-purpose system responsible for actively maintaining task-relevant information in the face of external distractions (24-26). Individuals high in working memory capacity (WMC) are not only better able to maintain information in the focus of attention, but they can also more efficiently retrieve information that has been momentarily displaced due to disruption (27-29). WMC has been implicated in task performance after an interruption (30), and the efficiency with which individuals multi-task (31). This suggests that WMC might influence the work strategies employed by ED doctors.

A number of studies have examined the role of workload in work management strategies. A study of medical and surgical doctors found that self-reported workload was significantly associated with the interruption rate (32). Another study from a Canadian hospital also found a strong association between interruption rates and department level workload measured as the average time from patient registration to doctor assessment (33). Both reported associations at an aggregated level, but were not able to provide insight into the possible direction of association or the mechanisms by which workload affects the prompt-response process or vice versa. There is also evidence to suggest that people opt for less cognitively demanding strategies when under load (34, 35). Thus it is not clear whether workload is associated with the overall number of prompts, whether it modifies the way clinicians use strategies, or some combination of both. Since both workload (36) and task-switching (11) have been associated with error, examining the relationship between these factors is a necessary part of understanding causes of error and will allow changes to clinical work to appropriately targeted.

2. Study objectives

2.1 Primary Objectives

1. To determine how the type of prompt affects which work management strategy a clinician uses in response.

2. To assess the relationship between workload and the use of work management strategies.

3. To examine the aspects of work that may contribute to the occurrence of memory lapses.

4. To compare the rate of prescribing errors and time to analgesia in the presence vs. absence of external prompts.

5. To observe the influence of Working Memory Capacity and polychronicity on the selection of work management strategies.

6. To determine the effect of workload in modifying the relationship between Working Memory Capacity and work management strategies

2.2 Secondary Objectives

There is much scope for more sophisticated quantitative analysis techniques in this field and developing such techniques is an important way to further our understanding of the complex systems under study. Hence a secondary aim of the study is:

1. To use the collected data to develop new quantitative approaches for analysing observational time-and-motion data.

3. Study Design

3.1 Design

This is an observational time-and-motion study in which features of the work of ED doctors will be recorded by observers using the Work Observation Method by Activity Timing (WOMBAT) system. This approach has been used in a number of studies to date to examine various aspects of clinical work practices (10-12, 37). The WOMBAT system allows the user to define a customised set of dimensions (e.g. task type) and categories within each dimension (e.g. direct care, documentation, etc.). The dimensions and categories designed specifically for this study are detailed in section 3.4 below. Observers will shadow doctors and record time stamped information about tasks, prompts and response strategies.

Observations will be carried out on doctors working in the acute section of the ED ~~during day shifts~~. The focus on the acute setting rather than fast-track is more likely to capture variations in work intensity. Observation sessions will cover the last two hours of a shift, including the completion of the shift, to enable the identification of non-resumed tasks or any action that is deferred but not followed up by shift's end.

Where prescribing is observed, those medication charts will be reviewed at a later time for any prescribing errors and analgesic prescriptions. Short, one-off, computer-based tests of working memory capacity and polychronicity (the propensity to multitask) will be administered at a time convenient to the participant.

3.2 Setting and Participants

This study is intended to be conducted in the Prince of Wales Hospital ED. Observations will be carried out on ED doctors who have agreed to participate. The only inclusion criteria are for participants to be working as residents, registrars or consultants/staff specialists in the ED at the time the study is carried out, are willing to give written informed consent and are willing to participate in and comply with the study. Interns will be excluded from the study.

3.3 Duration

Data collection is proposed to take place between May and October 2014. Analysis and preparation of manuscripts for peer-reviewed journals is then intended to commence by October 2014 and to be completed by December 2015.

3.4 Data collection

3.4.1 Observation with WOMBAT system

The WOMBAT system allows data collection fields to be tailored to the particular study prior to commencing observations. Essentially any number of categorical variables, or *dimensions*, can be predefined, each comprising any number of categories. The dimensions and categories in Table 1 build upon those used in previous studies, while being customised to the aims of this study. They have been chosen to be clinically relevant as well as being easily identified (i.e. with minimal ambiguity) by observers when observing doctors' work. The frequency with which tasks of a particular type were observed to occur in a previous ED-based study has informed the grouping of categories and sub-categories in this study so that the more frequent task types require the least number of screen 'clicks' to record.

Table 1. Dimensions, categories and sub-categories to be used for data collection in the WOMBAT system.

Dimension	Category	Sub-category
Task type*	Direct care	
	Indirect care	
	Documentation	
	Clinical communication	
	Management communication	
	Social communication	
	Prompt – advice	
	Prompt – action	
	Prompt – other communication	
	Prompt - other	Phone
		Equipment alarm
		MET call
	Prescribing	
	Other	Other communication
		Supervision
		On break
With Whom*	Consultant	
	Registrar	
	Resident/Intern	
	Nurse	
	Patient	
	Relative	
	Ambo	
	Other	Pharmacy
		Allied Health
		Admin
		Security/Police
		Other
	No one	
With what	Mobile	
	Landline	
	Computer	
	Paper	
Med chart sections	Once only	Line 1
		Line 2
		Line 3
		Line 4
		Line 5
	PRN	Line 1
		Line 2
		Line 3
		Line 4
		Line 5
		Line 6
		Line 7
	Regular 1	Line 1
		Line 2
		Line 3

	Regular 2 Other Fluid form Insulin chart Heparin infusion chart PCA form	Line 1 Line 2 Line 3 Line 4 Line 5 Line 6 Variable dose VTE Warfarin
Action	Deferral Deflection	

* Compulsory dimensions

Rather than include an explicit dimension that records the strategy used in response to a prompt, the design is such that strategies can be determined from the data during the analysis phase. A prompt will be recorded as a task with a particular duration (prompt type shown in table 1). The response strategy will be indicated by what precedes and follows the prompting task as follows:

- Deferral – the task following the prompt is the same as the task preceding the prompt *and* the prompt will be recorded as deferred (see section 3.4.2).
- Deflection – the task following the prompt is the same as the task preceding the prompt and the prompt will be recorded as deflected.
- Multitasking – more tasks occur in parallel after the prompt than before.
- Task-switching – the task following the prompt is different to the one preceding it.

3.4.2 Work related error

One observable way in which prompts may contribute to error is by causing memory lapses. For example, suspending a task in order to attend to something else increases the chance that that task is resumed at the wrong place in a sequence of steps or that it is not resumed at all (38). Along similar lines we plan to record instances in which action is deferred and whether the promised action is then followed up or not. Follow up may include handing the task over to someone else subsequent to the deferral. It is possible to record the deferral in the WOMBAT system, however, it is not practical to also record the follow-up status in a way that connects it to the original deferral. These will be therefore concurrently recorded by hand with sufficient detail so that completion of the promised action at some later time can be identified. Conducting observation sessions so that they encompass the end of the doctor's shift means that any deferred action that has not been followed up can be determined with certainty. Since some deferred tasks may have been followed-up in a way that was not apparent to the observer, post-shift questioning of the observed doctor will ascertain whether outstanding tasks were actually followed up (see question 5 in Appendix 1). This will also remind the doctor to any forgotten tasks that need to be followed up.

Another type of memory lapse that is easily ascertained from the data is non-resumption of a task that is suspended when a clinician switches to another task. Such tasks are recorded by the software as still pending by the end of the observation session. Conducting each observation session until the end of the shift is also necessary for this kind of error to ensure that any tasks recorded as suspended could not have been resumed after the cessation of observation.

3.4.3 Prescribing error and time to analgesia

Observable outcomes related to clinical practice are the occurrence of error in the prescribing of medications and time to analgesia. There is potential for prompts during prescribing to increase the risk of error and for increased prompts frequency during work in general to slow down the prescribing of a patient's pain management. When prescribing is observed, this will be recorded in the WOMBAT system as a prescribing task (see Table 1). At the same time, details of which sections of the national inpatient medication chart were completed during that task will also be recorded with WOMBAT. Whether an "Adult fluid form", "Patient controlled analgesia form", "Heparin infusion chart" or "Subcutaneous insulin chart" was filled in will also be recorded with WOMBAT. The patient MRN on the chart will be noted by hand on a separate sheet. The MRNs for subsequent prescribing tasks will be noted in consecutive order on the sheet during an observation session.

At the end of each observation session, data will be uploaded from the WOMBAT system and an identifier for each prescribing task will be assigned. This identifier will also be recorded on the sheet containing the MRNs. The prescribing tasks in the WOMBAT data can be matched to the MRNs by the fact that both appear in time order within the observation session. The sections of the medication charts where prescribing was observed will be reviewed for errors and analgesic prescriptions at regular intervals.

The review of medication charts will be carried out by a hospital ED pharmacist. The hospital pharmacist will be provided with a list of MRNs and medication orders to review at regular intervals. The medication charts will be requested from medical records. The hospital pharmacist will also refer to the relevant patient clinical notes in the electronic ED patient records. Hospital pharmacists regularly access patient clinical notes in the course of their work to assess medication appropriateness. For example, the pharmacist needs to check a patient's renal function in order to check that a medication dose is therapeutic and not toxic.

Errors will be categorised according to the error classifications shown in Appendix 2, which has been used by the investigators in an earlier published multi-centre prescribing error study in Sydney (17). The patient's year of birth will be noted from the records so that patient age can be adjusted for in the analysis since older patients have more complicated medication regimens and may be more at risk of clinical prescribing errors. Additionally, the time of presentation to the ED and the time at which analgesics were administered will be noted so that the time to analgesia variable can be calculated. Details of the reviews will initially be recorded on paper format and subsequently transferred and stored in an electronic file using only the study-specific identifier, not the MRN. This information will then be matched (via the study-specific identifier) with the prescribing tasks in the WOMBAT data in preparation for analysis. The sheets on which the MRNs were written will be stored in a locked cabinet at UNSW inside a security monitored building. This paper-based information will be destroyed after the study completion. No patient identifiers other than the MRN and year of birth will be recorded and no MRNs will be stored in electronic form at any time.

3.4.4 Workload

Workload has been shown to affect clinicians' work including the use of work management strategies. Workload has been associated with the rate of task-switching in two different studies, one of which was conducted in an ED (32, 33). Also it has been suggested that task-switching is favoured over multitasking in situations of high workload (34) and that high workload can increase the risk of error (36). Hence it is important to adjust for the effects of workload in the assessment of the primary study objectives.

The previous study of work in the ED used the number of daily presentations as a measure of work load (10). However, this did not appear to be associated with any aspects of the work process. Measures of work intensity at a higher temporal resolution may be required. The first proposed measure in this study is the time-specific number of patients in the ED system stratified by triage category. This is a measure of departmental load and improves on the number of daily presentations in that it takes into account temporal variation and includes a measure of case severity. This will be obtained by extracting data from the ED's patient administration system after the completion of the observation phase. The data will be patient-level (i.e. one record per patient visit), but only information about time and date of arrival, time of discharge/admission, and triage category will be required. MRN is not necessary for this data and will be replaced by an alternative identifier that, once assigned, cannot be used to reconstruct the MRN. Hence the researchers will have no access to the MRNs on this data at any point. Any reporting on this data will only be in via summary statistics and aggregated counts. Data relating to each session will be extracted and merged with the WOMBAT data at the commencement of the analysis so that the time specific workload measures can be synchronised with the dates and times of data collection sessions.

An even higher resolution measure of workload is the observed clinician's time-specific heart rate variability (HRV). This has been shown to be an indicator of workload in comparable settings (39-42). Participating doctors will be required to wear a wrist watch style heart rate monitor and a sensor mounted on a chest strap (Polar RS800CX) for the time that they are under observation. This will record a time stamp for each heart beat from which an estimate of HRV at any given time can subsequently be calculated. Other measures have been used for the same purpose, however, HRV is less affected by general exertion than heart rate and is easier to measure continuously than both galvanic skin response and blood pressure.

3.4.5 Working memory capacity and polychronicity

Both working memory capacity (WMC) and polychronicity are potential individual-level factors that may influence the way an individual manages external prompts. WMC will be assessing using the procedure recommended by Unsworth, Heitz, Schrock and Engle (43). The battery consists of one computer-administered task taking approximately twenty minutes. The test is called the Automated Operation Span Task, and is administered using the Inquisit Lab Software (44). On each trial, participants see an alternating sequence of arithmetic equations (e.g., 3 + 2 = 5) and to-be-remembered consonants. The participants have to judge the correctness of the equation and encode the following consonant for subsequent recall. The trials consist of three sets of each list size, with the list sizes ranging from 3 to 7. This makes for a total of 75 letters and 75 sums. Recall on each trial consists of clicking the box next to the appropriate letters in the correct orders. Recall is untimed. This version of the Operation Span test correlates well with other measures of WMC, and has both good internal consistency ($\alpha = .78$) and test-retest reliability (.83) (43). Since WMC changes with age it will be necessary to record the age of each participant in years.

Polychronicity is defined as the extent to which people prefer to be engaged in two or more tasks at a time and believe that their preference is the best way to complete tasks (45). This preference has been related to dimensions of personality (46), performance under pressure, changing deadlines (47), and a sense of time urgency (48). Specifically, individuals who are polychrons are likely to demonstrate higher levels of multi-tasking behaviours than monochrons.

Polychronicity will be assessed using an adapted version of the Inventory of Polychronic Values (IPV) [Appendix 4] (45). Participants respond to ten items rated on a seven-point Likert scale ranging from 'Strongly Agree' (7) to 'Strongly Disagree' (1). Five of the ten items are reverse-scored to account for acquiescence bias. An example item from the scale is "I like to juggle several activities at the same time". The median internal reliability for the scale across nine samples was 0.84, with the test-retest reliability estimates ranging between 0.78 and 0.95.

3.4.6 Other performance related factors

Fatigue is associated with performance (49) and is potentially associated with a clinician's choice of strategy for a given prompt. Hence it is considered an important factor to adjust for in the analysis phase. As all observation sessions will cover the last two hours of the day shift, there will be no variation in fatigue levels due to time of shift. However, there may be variation between doctors and between observation sessions for the same doctor. An assessment of fatigue corresponding to each observation session will be collected via a series of questions asked at the end of each observed shift (see Appendix 1). While considerable research has gone into the measurement of fatigue in motor vehicle drivers, there is no clear consensus about the best continuous measures of fatigue and those in use could not be implemented in the ED setting without impairing doctors' ability to work.

4. Study outline

4.1 Investigation plan

1. Obtain ethics approval from both PoW and UNSW HRECs

2. Conduct information sessions for potential participants

3. Obtain consent and enrol those doctors who agree to participate

4. Practice use of the WOMBAT tool with the dimensions and categories listed above to ensure accuracy and internal consistency for the actual data collection. In this phase the physiological monitoring device will also be tested in conjunction with WOMBAT observations to ensure accurate time synchronisation of the two data recording processes.

5. Carry out observation sessions on participating doctors and perform one-off working memory capacity and polychronicity tests. Periodically review medication charts during this period.

6. Data cleaning, analysis and preparation of publications

4.2 Study procedure risks

It is possible that being observed may cause some doctors to feel uncomfortable. This will be addressed by providing comprehensive information about the study prior to obtaining consent and by making it clear that participants may withdraw at any time, and having withdrawn may request their data to also be withdrawn. Another possible risk is that individuals may be identified with particular memory lapses. This will be mitigated by only recording participants in the data via their study ID number. Also both the data and any documentation linking ID numbers with other personal information will be kept on password protected servers at UNSW. Any publications will only report data aggregated in such a way that re-identification of an individual is not possible.

Wearing a heart rate monitor may be unfamiliar or distracting for some participants. If that is the case they may withdraw from this component of the study at any time while still participating in observation sessions.

This study will review medication charts in order to identify prescribing errors. Thus, a situation may occur where a prescribing error is identified which has the potential to cause patient harm. This may be immediately after prescribing or at a later stage when the assessment is carried out. If this happens, the pharmacist will follow the “serious error protocol” which is in line with the approach used by hospital pharmacists during their normal work day (Appendix 3).

In order to examine patient medication charts and clinical notes for identification of prescribing errors at a time subsequent to the prescribing event, patient MRNs and observed chart sections that were completed will be recorded during observations. Aside from MRNs, only the year of birth will be recorded. The MRNs will not be kept in electronic form, will not appear on the main dataset and will not be reported in any of the research outputs. Paper copes of the MRNs will be kept in a locked cabinet inside a secure building at UNSW and will be destroyed after study completion

There is also a risk that the prescribing errors identified may be linked to specific doctors. In order to reduce this risk, doctors will be assigned a unique identifier during the data collection process. Thus, the final datasets will not contain doctors’ names, and the list linking names with identifiers will be stored separately from the main data. Doctors will not be able to be linked to any errors in any published outputs.

It is also possible that doctors may feel uncomfortable being observed. This will be minimized through regular visits to the ED so that the participating doctors are familiar with the researchers and study process. Previous experience in similar studies has shown that doctors are comfortable being observed.

Working Memory Capacity tests are sometimes experienced as difficult by research participants. Items in the test are timed and as such participants only have a limited window of opportunity to provide answers. To mitigate unfamiliarity with WMC tests, participants who agree to complete the test will be provided with two-three practice items before they commence the task. Participants will also be informed that WMC tests are meant to be difficult to complete all items correctly in the time allotted, and that they should not be concerned if they did not complete all items on the test. This will be included on the information statement and reiterated to the participant before they commence the test.

From previous experience, we know that participants often enjoy the opportunity to complete intelligence tests of this nature, and also enjoy receiving feedback from these tests. If participants request, an individualised report of their WMC test result can be provided to them. They may also request a summary of their heart rate monitoring data and work observation data (see Participant Feedback Form – version 1). Feed on test results of heart rate monitoring will not be provided to anyone other than the individual to whom it relates, and even then only on request.

4.3 Recruitment

In consultation with a senior ED physician at Prince of Wales, information sessions will be organised where the details of the study will be described to potential participants and copies of the Information for Participants sheet and Consent form will be provided. Doctors not able to attend information sessions will be approached individually.

4.4 Informed consent process

Participants will be given information about the study prior to agreeing to participate (see Information for Participants sheet), will give written consent if they agree to participate, and will have the option to withdraw at any time without justification (see Consent Form). No coercion, rewards or incentives will be involved in an individual's decision to participate in the study.

4.5 Enrolment procedure

The participant will be enrolled into the study after the informed consent process has been completed and the participant has met the inclusion criteria. The participant will receive a study enrolment number and this will be used to identify doctors during data collection.

4.6 Patient consent

Although patients are not the focus of this study, observed doctors will be instructed to briefly introduce observers when possible, to make it clear that observations are focused on the doctor and to ask the patient if they consent to the observations continuing during their treatment. The recording of MRNs will only occur for the purpose of reviewing medication charts. MRNs will not be recorded in the final electronic datasets and paper forms containing MRNs will be destroyed at the end of the study.

5. Safety

5.1 Adverse event reporting

An adverse event is defined as any untoward or unfavourable occurrence to a participant related to the study. The risk of such an event as a direct cause of this study is highly unlikely and no such event has occurred during the many similarly designed studies carried out by our research centre. Should any adverse events occur, the researcher will report the event to the HREC as required by the National Statement on Ethical Conduct in Research Involving Humans.

5.2 Confidentiality

Data will be recorded in a de-identified way. Any information that can be linked to an individual either directly or through re-identification, will be kept on password protected secure servers and will only be available to approved study investigators. Results from this study are intended to be published in a series of peer-reviewed scientific articles. These publications will not identify the site nor individual participants and will only report data on work practices in an aggregated form. Information that links participant ID with personal identifiers will be permanently deleted as soon as the data collection phase is complete.

5.3 Conduct during observations

Some patient treatment in the ED takes place behind curtains or in consulting rooms. In the instance that a doctor feels the observer's presence to be inappropriate, they may ask the observer to remain outside the curtains or consulting room. If possible, observers will continue recording outside the curtain or room based on auditory cues, otherwise they will record the period as "direct care". However, if a patient asks that their treating doctor not be observed or the doctor asks that no information be recorded while treating a particular patient, then all observations will be suspended until the doctor moves to a task unrelated to that patient. If any doctors, observed or otherwise, are performing resuscitation, observers will stay sufficiently far from the resuscitation area so as not to impede the resuscitation in any way. As a general rule, observers will minimise their impact on both staff and patients as much as possible at all times.

6. Statistical Considerations

6.1 Sample size considerations

Due to the observational nature of this study and the complexity of the socio-technical context of the ED, traditional sample size calculations to determine a required amount of observation time are too simplistic for primary objectives 1 and 2. Various approaches have been suggested to calculate sample sizes for multivariate analyses [see for example (50)], however, their complexity and reliance on many assumptions makes them difficult to apply meaningfully in non-experimental settings. Previous studies carried out using the WOMBAT system in settings similar to this study have produced detailed and clinically relevant results that have been published in peer-reviewed journals based on between 100 and 500 hours of observation time (10-12, 37, 51). The volume of observation time in other clinical observational studies ranges from around 30 hours up to several hundred. Hence to meet the study objectives within time and resource constraints we propose to carry out 100 hours of observations over approximately 50 sessions. Based on the frequency of task-switching and multitasking reported in previous studies, coupled with our intention to record two additional strategies (deferral and deflection), it is expected that the proposed 100 hours of observations will yield sufficiently many instances of strategy use to allow us to comprehensively address objectives 1 and 2. The sessions will be randomly distributed across the participating doctors.

In terms of assessing the effect of prompts on prescribing error risk, 100 hours of observation would enable detection of an error proportion of 0.22 in the presence of prompts compared to 0.1 without prompts. This is based on a test for a difference between two independent proportions assuming 80% power, 5% significance, equally sized groups and three prescriptions per hour (as observed in piloting). A baseline error rate of 0.1 is conservative in comparison to reported values in the literature and as this value approaches 0.5, smaller effects will be able to be detected with the same sample size. Hence 100 hours of observation will be considered a lower bound.

For objective 3, memory lapses are expected to be relatively infrequent. In data from a previous ED-based study there was roughly one non-resumed task every 1 hour and 40 minutes (10). Another study of interruptions during CPOE system use observed an uncompleted task or memory lapse every 45 minutes on average (52). Due to the anticipated small sample of deferred actions that aren't followed up, the analysis related to objective 3 will be largely descriptive.

6.2 Statistical analysis plan

1. In order to assess the effect of prompts on the use of work management strategies, multinomial logistic regression will be applied as a task-level analysis. The outcome will be the strategy (task-switch, multitask, defer or deflect) and the covariates will comprise the prompt type, and primary and secondary task types associated with each recorded strategy. The regression model will also adjust for the effect of workload, working memory capacity and fatigue. A random intercepts approach will be used to account for correlation within tasks performed by the same individual clinician.

2. The role of workload in the prompt-response process will be analysed by examining associations in two directions. First we will examine whether HRV during a short window immediately preceding a prompt is associated with the strategy use, after adjusting for prompt type. This is to assess whether workload influences the way clinicians handle prompts. We will then look at whether workload, as measured by HRV, in a short window immediately following a prompt is associated with prompt type to determine whether prompts contribute to a clinician's workload.

3. While a multivariate analysis of factors associated with memory lapses would be preferable, the proposed descriptive analysis (due to small sample size) will still provide valuable insights into the mechanisms by which particular aspects of work may cause memory lapses. The type of task that is not resumed or followed up, the departmental and individual workload at the time of lapse and during the rest of the shift will be summarised by counts and proportions.

4. To determine the impact of prompts on prescribing error, logistic regression will be applied to all prescribing tasks. The error status of each task will be treated as the binary outcome of the model and the number of prompts will be the main covariate of interest. This relationship will be adjusted for confounding factors including prompt type, workload, prompt intensity during a time window preceding the prescribing task, and working memory capacity of the prescribing doctor.

5. To assess the effect of WMC and polychronicity on the propensity to choose particular strategies, measures of both factors will be included in the model described at point 1 above.

6. To see how workload can modify the relationship between WMC and work management strategies, the model described in point 1 will be re-run, but with stratification by quintiles of workload. This will be done separately for individual level workload (as measured by HRV) and for departmental workload.

Appendix 1: Post-observation questions

1. a) In the last 24 hours how many hours of sleep have you had
   b) Would you consider this amount of sleep to be above average, below average or about average for you on a workday?

2. In the past week how many shifts have you worked?

3. How many of those were night shifts?

4. When was the most recent night shift that you worked?

5. Tasks A, B, C, etc., were not observed to be followed up after having being deferred. Can you comment on the outcome of each?

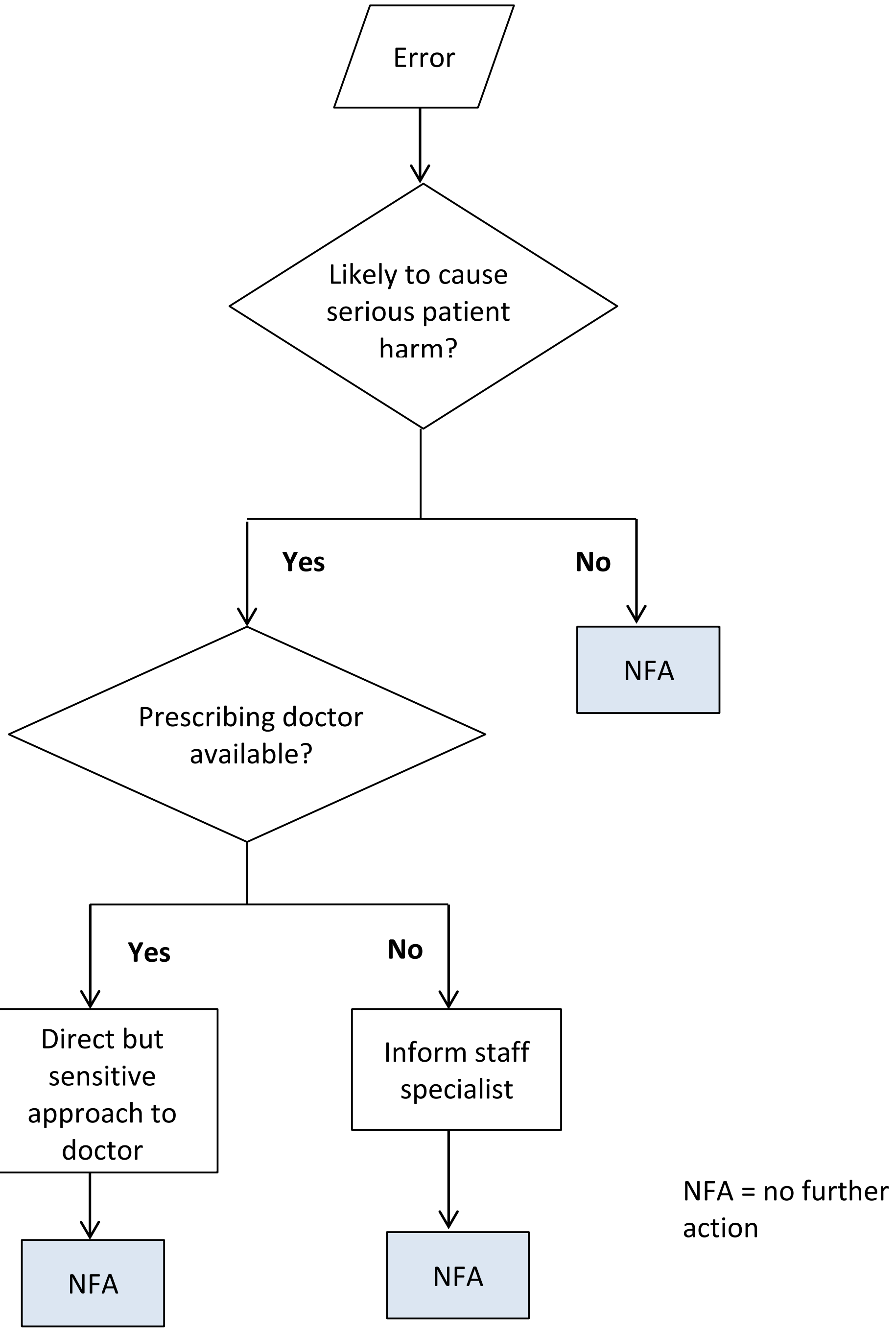
Appendix 2: Classification of prescribing errors (17)

CLINICAL ERRORS		
Error category	**Definition**	**Examples**
Wrong drug	Occurs when an inappropriate medication or IV fluid is prescribed e.g. the drug prescribed is not indicated for the patient's condition; the drug or IV fluid is contraindicated for a coexisting condition; or an IV drug is prescribed with an incompatible diluent Note: Excludes generic substitution	e.g. hydrocortisone 25mg oral mane was prescribed instead of cortisone; chamomile lotion was ordered instead of calamine lotion
Wrong dose / volume	Occurs when the prescribed medication dose or IV fluid volume is higher or lower than that recommended for the condition, taking into account the patient's age, weight, renal and liver function May also occur when a dose is not altered in response to abnormal drug serum levels or laboratory tests Note: A dose may differ from normal recommended reference ranges and not be classed as an error where it is accepted practice to do so, i.e. the dose may have been queried by a pharmacist, but the specialist physician insisted on the prescribed dose, e.g. high dose flucloxacillin despite severe renal impairment in patients with severe infection when recommended by the infectious diseases team; low doses of tricyclic antidepressants initiated by the pain team.	
Wrong rate / frequency	Occurs when the prescribed frequency of administration of a drug or an IV rate falls outside the recommended range	
Wrong route	Occurs when a medication is prescribed via an incorrect route of administration	e.g. IV medication was prescribed orally; left eye was written instead of right eye
Wrong formulation	Occurs when the wrong dosage form of a medication is ordered	e.g. an immediate release tablet was prescribed when an extended release form was required
Wrong timing	Occurs when a drug is prescribed at the wrong time of day	e.g. simvastatin prescribed in the morning instead of the evening (it is more efficacious when taken at night)
Wrong strength	Occurs when the prescribed drug strength is incorrect; the concentration of an IV infusion is prescribed incorrectly; or a dose is prescribed that does not exist or would not be able to be obtained easily from the current dose forms	e.g. mg was prescribed instead of micrograms (or vice versa) e.g. alendronate 75mg tab oral, take one tab once weekly (weekly dose only available as 70mg tablets)
Wrong patient	Occurs when a medication is prescribed for the wrong patient e.g. the prescriber writes a drug order intended for patient A on the medication chart belonging to patient B	
Not prescribed	Occurs when a medication clinically indicated	

	for the patient is not prescribed; or the drug is not reordered when the patient's medications are recharted	
Not indicated	Occurs when a drug which is not indicated is prescribed for the patient; a drug is continued following a clinically significant adverse drug reaction; a drug which is no longer indicated is reordered; or a drug which should have been discontinued has not been ceased May also occur when a prescriber fails to cease/withhold a drug in response to abnormal drug serum levels or laboratory tests	e.g. fluticasone/ salmeterol inhaler prescribed for a patient without chronic obstructive airways disease e.g. an antibiotic which was not discontinued after completion of the course
Duplicated drug therapy	Occurs when two orders have been prescribed for one medication and both orders are active; there are two active orders for the same medication on two different charts; or the same drug is prescribed twice, as a single agent and as a combination product May also occur when two drugs are prescribed for the same indication when only one is necessary	e.g. one order was prescribed by generic and one by brand name e.g. ranitidine and omeprazole for gastro-oesophageal reflux disease
Drug-drug interaction	Occurs if two of the drugs prescribed for a patient are known to have a clinically significant interaction and this interaction is not acknowledged and monitored	
Allergy	Occurs when a drug is prescribed for a patient with a documented clinically significant allergy to that drug/class of drugs	
Inadequate monitoring	Occurs when the prescriber fails to order appropriate and timely clinical or laboratory tests to assess the patient's response to prescribed therapy Note: if adequate lab tests are ordered, but the results are not acted upon accordingly, resulting in potential or actual compromised patient care, this may be classed as wrong dose/volume error	
PROCEDURAL ERRORS		
Unclear order	Occurs when the prescription is unclear or ambiguous e.g. the writing is illegible; or the order contains additional comments which apparently contradict the medication order	e.g. clotrimoxazole topical interdigital BD (the prescriber was confused between cotrimoxazole and clotrimazole)
Incomplete order	Occurs when the order does not include all the necessary information i.e. drug name; strength (if appropriate); formulation (if appropriate); dose; route of administration; frequency ; the diluent for injectables; duration of time and/or rate of infusion (IV infusions); duration of time (IV fluids)	
Legal/Procedural	Occurs when an aspect related to the prescription does not comply with the law, the NSW Department of Health or the hospital policy (and has not been assigned as an unclear order); the allergy field of the medication chart has not been completed; or the strength, dose, route or frequency of an existing handwritten medication order has been altered (such a change legally requires the entire order to be recharted)	

Appendix 3: Serious error protocol

When an UNSW researcher identifies a prescribing error, they will first need to decide if the error can lead to serious patient harm. They will then follow the decision tree below.



Examples of prescribing errors that may cause serious harm:

- Prescribing a medication to a patient who has a history of allergy or adverse drug reaction to that medication.
- Prescribing an overdose of a medication e.g. methotrexate prescribed daily instead of weekly.

Examples of prescribing errors that would not be considered to cause serious harm:

- Procedural errors, such as the absence of a signature.
- Prescribing a medication dose using an incorrect abbreviation or unit e.g. writing U instead of units or prescribing a liquid dose in ml instead of mg or g.

Appendix 4

POLYCHRONIC VALUES INVENTORY

This questionnaire measures your preference for doing multiple things at once, versus completing one task at a time. Please answer each statement below by putting a circle around the number that best reflects your degree of agreement or disagreement with that statement. Use the seven point response scale below.

1 Strongly Disagree	2 Moderately Disagree	3 Slightly Disagree	4 Neither Agree nor Disagree	5 Slightly Agree	6 Moderately Agree	7 Strongly Agree

1	I like to juggle several activities at the same time.	1	2	3	4	5	6	7
2	I would rather complete an entire project every day than complete parts of several projects.	1	2	3	4	5	6	7
3	I believe people should try to do many things at once.	1	2	3	4	5	6	7
4	When I work by myself, I usually work on one project at a time.	1	2	3	4	5	6	7
5	I prefer to do one thing at a time.	1	2	3	4	5	6	7
6	I believe people do their best work when they have many tasks to complete.	1	2	3	4	5	6	7
7	I believe it is best to complete one task before beginning another.	1	2	3	4	5	6	7
8	I believe it is best for people to be given several tasks and assignments to perform.	1	2	3	4	5	6	7
9	I seldom like to work on more than a single task or assignment at the same time.	1	2	3	4	5	6	7
10	I would rather complete parts of several projects every day than complete an entire project.	1	2	3	4	5	6	7

References

1. Speier C, Valacich JS, Vessey I. The influence of task interruption on individual decision making: An information overload perspective. Decision Sciences. 1999;30(2):337-60.
2. Altmann EM, Trafton JG, editors. Task interruption: Resumption lag and the role of cues. 26th Annual Meeting of the Cognitive Science Society; 2004; Chicago.
3. Bailey BP, Konstan JA. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. Comput Human Behav. 2006;22(4):685-708.
4. Hopkinson SG, Mowinski Jennings B. Interruptions during nurses' work: A state-of-the-science review. Research in Nursing and Health. 2013;36:38-53.
5. Colligan L, Bass EJ. Interruption handling strategies during paediatric medication administration. BMJ Quality and Safety. 2012;21(11):912-7.
6. Brixey JJ, Robinson DJ, Johnson CW, Johnson TR, Turley JP, Zhang J. A concept analysis of the phenomenon interruption. Advances in Nursing Science. 2007;30(1):E26.
7. Liu D, Grundgeiger T, Sanderson PM, Jenkins SA, Leane TA. Interruptions and blood transfusion checks: lessons from the simulated operating room. Anesthesia & Analgesia. 2009;108(1):219-22.
8. Grundgeiger T, Sanderson P, MacDougall HG, Venkatesh B. Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. J Exp Psychol Appl. 2010;16(4):317-44.
9. Walter SR, Li L, Dunsmuir WTM, Westbrook JI. Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours
BMJ Qual Saf. 2014;23(3):231-41.
10. Westbrook JI, Coiera E, Dunsmuir WTM, Brown BM, Kelk N, Paoloni R, et al. The impact of interruptions on clinical task completion. Qual Saf Health Care. 2010;19(4):284-9.
11. Westbrook JI, Woods A, Rob MI, Dunsmuir W, Day RO. Association of interruptions with an increased risk and severity of medication administration errors. Arch Intern Med. 2010;170(8):683.
12. Arabadzhiyska P, Melissa B, Walter SR, Day RO, Westbrook JI. Shedding light on junior doctors' work practices after hours. Internal Medicine Journal. 2013;in press. Epub 26 June 2013.
13. Baethge A, Rigotti T. Interruptions to workflow: Their relationship with irritation and satisfaction with performance, and the mediating roles of time pressure and mental demands. Work & Stress. 2013;27(1):46-63.
14. Lewis PJ, Dornan T, Taylor D, Tully MP, Wass V, Ashcroft DM. Prevalence, incidence and nature of prescribing errors in hospital inpatients: a systematic review. Drug safety : an international journal of medical toxicology and drug experience. 2009;32(5):379-89. PubMed PMID: 19419233. Epub 2009/05/08. eng.
15. Ryan C, Ross S, Davey P, Duncan EM, Francis JJ, Fielding S, et al. Prevalence and causes of prescribing errors: the PRescribing Outcomes for Trainee Doctors Engaged in Clinical Training (PROTECT) study. PloS one. 2014;9(1):e79802. PubMed PMID: 24404122. Pubmed Central PMCID: PMC3880263. Epub 2014/01/10. eng.
16. Tully MP, Ashcroft DM, Dornan T, Lewis PJ, Taylor D, Wass V. The causes of and factors associated with prescribing errors in hospital inpatients: a systematic review. Drug safety : an international journal of medical toxicology and drug experience. 2009;32(10):819-36. PubMed PMID: 19722726. Epub 2009/09/03. eng.
17. Westbrook JI, Reckmanm M, Li L, Runciman WB, Burke R, Lo C, et al. Effects of two commercial electronic prescribing systems on prescribing error rates in hospital in-patients: a before and after study. PLoS Medicine. 2012;9(1):e1001164.
18. Rosini JM, Grovola MR, Levine BJ, Jasani NB. Prescribing habits of vancomycin in the Emergency Department: are we dosing appropriately? The Journal of emergency medicine. 2013 May;44(5):979-84. PubMed PMID: 23466019. Epub 2013/03/08. eng.

19. National Research Council. To Err Is Human: Building a Safer Health System. Washington, DC: The National Academies Press; 2000.
20. Dormann H, Sonst A, Muller F, Vogler R, Patapovas A, Pfistermeister B, et al. Adverse drug events in older patients admitted as an emergency: the role of potentially inappropriate medication in elderly people (PRISCUS). Deutsches Arzteblatt international. 2013 Mar;110(13):213-9. PubMed PMID: 23596501. Pubmed Central PMCID: PMC3627162. Epub 2013/04/19. eng.
21. Patanwala AE, Hays DP, Sanders AB, Erstad BL. Severity and probability of harm of medication errors intercepted by an emergency department pharmacist. The International journal of pharmacy practice. 2011 Oct;19(5):358-62. PubMed PMID: 21899616. Epub 2011/09/09. eng.
22. Reckmanm M, Li L, Westbrook JI. Protocol for Measuring Prescribing Errors Pre and Post Electronic Medication Management System Introduction. Kensington: Centre for Health Systems and Safety Research, University of New South Wales, 2011.
23. Lewis PJ, Ashcroft DM, Dornan T, Taylor D, Wass V, Tully MP. Exploring the causes of junior doctors' prescribing mistakes: a qualitative study. British journal of clinical pharmacology. 2014 Feb 11. PubMed PMID: 24517271. Epub 2014/02/13. Eng.
24. Baddeley A. WOrking memory, thought and action. Oxford, UK: Oxford University Press; 2007.
25. Engle RW, Kane MJ. Executive attention, working memory capacity, and a two-factor theory of cognitive control. The Psychology of Learning and Motivation. 2003;44:145-99.
26. Kane MJ, Bleckley MK, Conway ARJ, Engle RW. A Controlled-Attention View of Working-Memory Capacity. Journal of Experimental Psychology: General. 2001;130(2):169-83.
27. Brewer GA, Unsworth N. Individual differences in the effects of retrieval from long-term memory. Journal of Memory and Language. 2012 4//;66(3):407-15.
28. Unsworth N, Brewer GA. Examining the relationship among item recognition, source recognition, and recall from an individual differences perspective. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2009;35(6):1578-85.
29. Unsworth N, Engle RW. The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. Psychological Review. 2007;114(1):104-32.
30. Ball BH, Knight JB, Dewitt MR, Brewer GA. Individual differences in the delayed execution of prospective memories. The Quarterly Journal of Experimental Psychology. 2013;66(12):2411-25.
31. Bühner M, König CJ, Pick M, Krumm S. Predictors of the Speed and Error Aspect of Multitasking Performance. Human Performance. 2006;19(3):253-75.
32. Weigl M, Müller A, Vincent C, Angerer P, Sevdalis N. The association of workflow interruptions and hospital doctors' workload: a prospective observational study. BMJ Qual Saf. 2012;21(5):399-407.
33. Friedman S, Elinson R, Arenovich T. A study of emergency physician work and communication: a human factors approach. Israeli J Emerg Med. 2005;5:35-42.
34. Wickens CD. Multiple resources and performance prediction. Theor Issues Ergon Sci. 2002;3(2):159-77.
35. Beilock SL, DeCaro MS. From Poor Performance to Success Under Stress: Working Memory, Strategy Selection, and Mathematical Problem Solving Under Pressure. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2007;33(6):983-98.
36. Levin S, Aronsky D, Hemphill R, Han J, Slagle J, France DJ. Shifting Toward Balance: Measuring the Distribution of Workload Among Emergency Physician Teams. Annals of Emergenc Medicine. 2007;50(4):419-23.
37. Westbrook JI, Duffield C, Li L, Creswick NJ. How much time do nurses have for patients? a longitudinal study quantifying hospital nurses' patterns of task time distribution and interactions with health professionals. BMC Health Serv Res. 2011;11(319).
38. Grundgeiger T, Sanderson P. Interruptions in healthcare: theoretical views. Int J Med Inform. 2009;78(5):293-307.

39. Healey JA, Picard RW. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. IEEE Transactions on Intelligent Transportation Systems. 2005;6(2):156-66.
40. Mehler B, Reimer B, Coughlin JF. Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand From a Working Memory Task: An On-Road Study Across Three Age Groups. Human Factors. 2012;54(3):396-412.
41. Hjortskov N, Rissen D, Blangsted AK, Fallentin N, Lundberg U, Søgaard K. The effect of mental stress on heart rate variability and blood pressure during computer work. Eur J Appl Physiol. 2004;92:84-9.
42. Mark GJ, Voida S, Cardello AV. “A Pace Not Dictated by Electrons”: An Empirical Study of Work Without Email In: Konstan JA, editor. CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 5-10 May 2012; Austin, Texas2012.
43. Unsworth N, Heitz RR, Schrock JC, Engle RW, Unsworth N, Heitz RR, et al. An automated version of the operation span task. Behav Res Methods. 2005;37(3):498-505.
44. Inquisit 4.0.5.0 [Computer Software]. Seattle, WA: Millisecond Software; 2014.
45. Bluedorn AC, Kalliath TJ, Strube MJ, Martin GD, Bluedorn AC, Kalliath TJ, et al. Polychronicity and the inventory of polychronic values (IPV): the development of an instrument to measure a fundamental dimension of organizational culture.(Polychronicity). Journal of Managerial Psychology. 1999;14(3 4):205.
46. Conte JM, Jacobs RR. Validity evidence linking polychronicity and big five personality dimensions to absence, lateness, and supervisory performance ratings. Hum Perform. 2003;16(2):107-29.
47. Kaufman-Scarborough C, Lindquist JD, Kaufman - Scarborough C, Lindquist JD. Time management and polychronicity: comparisons, contrasts, and insights for the workplace.(Polychronicity). Journal of Managerial Psychology. 1999;14(3 4):288.
48. Conte JM, Rizzuto TE, Steiner DD. A construct-oriented analysis of individual-level polychronicity. MCB UP Ltd; 1999. p. 269-88.
49. Dawson D, Reid K. Fatigue, alcohol and performance impariment. Nature. 2002;388:235.
50. Demidenko E. Sample size determination for logistic regression revisited. Statistics in Medicine. 2007;26:3385-97.
51. Westbrook JI, Ampt A, Kearney L, Rob MI. All in a day's work: an observational study to quantify how and with whom doctors on hospital wards spend their time. Med J Aust. 2008;188(9):506.
52. Collins SA, Currie L, Patel VL, Bakken S, Cimino JJ, editors. Multitasking by clinicians in the context of CPOE and CIS use. Medinfo; 2007; Brisbane, Australia.

A.3 Participant information and consent forms

Recruitment of participants largely involved approaching individual doctors and giving them a brief description of the study and what participation would involve. If they we interested in participating we gave them an information sheet to read and keep that gave details of the study, the terms of their involvement, and the confidentiality of any data pertaining to them. If they agreed to participate we asked them to sign a consent form. Hence all participants provided consent that was both written and informed. The forms used in recruitment are shown below.

INFORMATION FOR PARTICIPANTS

STUDY TITLE: Clinicians' Strategies to Manage Work in Emergency Care Settings

INVESTIGATORS: ***Mr Scott Walter***, PhD Candidate and Statistician, Centre for Health Systems and Safety Research, UNSW; ***Professor Johanna Westbrook***, Director, Centre for Health Systems and Safety Research, UNSW; ***Professor William Dunsmuir***, School of Mathematics and Statistics, UNSW; ***Ms Magda Raban,*** Post-doctoral Research Fellow, Centre for Health Systems and Safety Research, UNSW; ***Dr Heather Douglas,*** Post-doctoral Research Fellow, Centre for Health Systems and Safety Research, UNSW; ***Dr John Mackenzie***, Consultant, Prince of Wales Emergency Department; ***Ms Dana Strumpman***, Senior Pharmacist Clinical Services, Prince of Wales Pharmacy Department.

INTRODUCTION

You are invited to take part in a research study into understanding emergency department (ED) clinicians' work processes. This information sheet will tell you about what is involved in the study and help you decide whether or not you wish to take part. Please read this information carefully. If there is anything you do not understand or if you feel you need more information about anything, please ask.

WHY HAVE I BEEN ASKED TO TAKE PART?

You have been invited to take part in this research because you are a resident, registrar, consultant or staff specialist in the Prince of Wales ED.

WHAT IS THE PURPOSE OF THIS RESEARCH?

The aim of this study is to examine the ways in which doctors manage work in the ED. In particular we aim to use direct observation to assess the strategies used to handle interruptions. We also aim to determine how aspects of work in the ED may contribute to errors.

WHAT WILL HAPPEN TO ME IF I TAKE PART?

If you decide to participate, a researcher will shadow you while observing your work activities for approximately the last two hours of a shift in the acute section for up to five shifts spread randomly over a few months. The observer will record and time your activities using an electronic tablet. At the end of each observed shift you will be asked a few questions about your sleep and work schedule in the recent past. Also any tasks that appeared not to be followed-up by the end of the observation session will be discussed. At the end of your first observed shift you will be asked to complete a short questionnaire on your preference for multi-tasking. The post shift questions are expected to take no more than five minutes. During the observation sessions you will be asked to wear a heart rate monitor consisting of a wrist-watch style training computer and a sensor mounted on a chest strap (Polar RS800CX). We will also ask you to complete a twenty-minute computer-administered working memory task at a time chosen by you. When seeing each new patient, whenever possible, you will need to seek their consent for the observer to be present and indicate that they may suspend observations at any time.

CAN I WITHDRAW FROM THE STUDY?

Taking part in any research is entirely voluntary. If you do decide to take part you can withdraw at any time without having to give a reason and without prejudice or consequence. If you do not wish to wear a heart rate monitor during observation session or complete the multitasking or working memory test, you may opt out of these at any time and still remain in the study as an observee. On very rare occasions the observer may need to ask you a question to clarify recording of a task. The interruption to your work is expected to be minimal. The researcher will suspend observations at any time that you request.

You might find the working memory test difficult to complete in the time limits. This is normal, and you should not be concerned. You will have an opportunity to practice the task you will be asked to do before recording of your results starts. The researcher will end the task at any time that you request.

Information that links any of your personal identifiers to the main study data will be permanently deleted immediately after the data collection phase has been completed. All data will be stored electronically on a secure drive at the University of NSW for seven years after which time it will be permanently destroyed. It is anticipated that you will not incur any additional costs if you participate in this study, nor will you receive any payment for participating.

WHAT ARE THE POSSIBLE BENEFITS OF TAKING PART?

While we intend that this research study furthers understanding of ED clinicians' work patterns and the potential causes of error, we cannot and do not guarantee or promise that you will receive any direct benefits from this study.

WHAT ARE THE RISKS OF TAKING PART?

You may feel uncomfortable having someone watch you closely as you work. All research observers will make every effort to put you at ease. Data will be recorded in a de-identified way, and will not be linked to your name on the main dataset.

CONFIDENTIALITY

Provided that you have signed the accompanying consent form, any information that is obtained in connection with this study that can be identified with you will remain confidential. We plan to discuss the results with this hospital. The study may also be reported at conferences and in journals. In any publication, information will be provided in such a way that you cannot be identified. We will provide you with any new information that could influence your decision to remain in the study.

FURTHER INFORMATION

When you have read this information, we will discuss it with you further and answer any questions you may have. If you would like to know more at any stage, please feel free to contact Mr Scott Walter (02 9385 0510; scott.walter@unsw.edu.au). This information sheet is for you to keep.

This study has been approved by the Human Research Ethics Committee of the South Eastern Sydney Local Health District, Northern Sector (ref: 13/310). If you have any concerns or complaints about the conduct of the research study, you may contact the Executive Officer of the Ethics Committee, on (02) 9382 3587.

Clinicians' Strategies to Manage Work in Emergency Care Settings

Participant Consent Form

I, ……………………………………………………………….……………………………………………………[name] of

………[ward/department]

have been invited to participate in the above named research study and have discussed the study with ……………………………….………………………………………………… [name of informant]

• I acknowledge that I have received and read the Participant Information Sheet and that the purpose and nature of this research, including any possible risks, have been explained to me.

• I understand that my participation in this study is entirely voluntary and I may withdraw at any stage without providing a reason.

• I understand that information about my work and my physiology while performing that work will be recorded in a de-identified way and will be treated as strictly confidential. I agree that research data gathered from the results of the study may be published, provided I cannot be identified.

• I understand that I will be required to seek verbal permission from patients for the observer to be present and to indicate that the patient may ask for observations to be suspended at any time.

• I understand that the research project will be carried out according to the principles of the National Health & Medical Research Council National Statement on Ethical Conduct in Research Involving Humans.

• I understand that if I have any questions relating to my participation in this research study I may contact Scott Walter on (02) 9385 0510 or at scott.walter@unsw.edu.au to discuss any concerns.

• I understand that if I have any questions about my rights as a research subject, or on other administrative matters, I may contact the SESLHD Human Research Ethics Committee on (02) 9382 3587 or EthicsNHN@sesiahs.health.nsw.gov.au. [PTO]

• I hereby freely agree to participate in this research study.

Name (Print):………

Signature:………………………………………………………………………………… Date: …………………………..

Name of Witness (Print):……………………………………………………………………………………………

Signature of Witness:…………………………………………………………… Date: …………………………..

Clinicians’ Strategies to Manage Work in Emergency Care Settings
Participant Consent Revocation Form

I hereby wish to WITHDRAW my consent to participate in the research proposal described above and understand that such withdrawal WILL NOT jeopardise my relationship with the University of New South Wales or participating hospitals.

I do/do not (*delete as appropriate*) give consent for data collected about me up to the time of withdrawal to be retained as part of this study.

Name
(Print):………

Signature:………………………………………………………………………………………… Date: ………………………………..

The section for Revocation of Consent should be forwarded to:

Scott Walter

Centre for Health Systems and Safety Research

Level 1 AGSM Building

University of New South Wales

SYDNEY NSW 2052

scott.walter@unsw.edu.au