

“Enslaved to the Trapped Data”: A Cognitive Work Analysis of Medical Systematic Reviews

Ian A. Knight

Data-Driven Algorithms, Systems and Design Group
University of Nottingham
Nottingham, UK
ian.knight1@nottingham.ac.uk

David F. Brailsford

School of Computer Science
University of Nottingham
Nottingham, UK
david.brailsford@nottingham.ac.uk

Max L. Wilson

Mixed Reality Lab
University of Nottingham
Nottingham, UK
max.wilson@nottingham.ac.uk

Natasa Milic-Frayling

Data-Driven Algorithms, Systems and Design Group
University of Nottingham
Nottingham, UK
natasa.milic-frayling@nottingham.ac.uk

ABSTRACT

Systematic reviews are a comprehensive and parameterised form of literature review, found in most disciplines, that involve exhaustive analyses and rigorous interpretation of prior literature. Performing systematic reviews, however, can involve repetitive and laborious work in order to reach reliable standards, especially in strictly regulated domains where the quality of acquired knowledge is of critical importance, like in evidence-based medicine. This makes tools providing computerised assistance or automation during the task attractive. However, it is unclear which aspects of this Work Task are best suited for such support. This paper describes a three-month ethnographic study and Cognitive Work Analysis of the systematic reviews performed by a medical research group. Our findings provide the basis for future work designing search tools with localised optimization and subtask automation to support specific phases of the process.

CCS CONCEPTS

• **Human-centered computing** → *Ethnographic studies*; • **Information systems** → Information extraction; Summarization.

KEYWORDS

Work task; Systematic review; Cognitive work analysis

ACM Reference Format:

Ian A. Knight, Max L. Wilson, David F. Brailsford, and Natasa Milic-Frayling. 2019. “Enslaved to the Trapped Data”: A Cognitive Work Analysis of Medical Systematic Reviews. In *2019 Conference on Human Information Interaction and Retrieval (CHIIR '19)*, March 10–14, 2019, Glasgow, United Kingdom. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3295750.3298937>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIIR '19, March 10–14, 2019, Glasgow, United Kingdom

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6025-8/19/03...\$15.00

<https://doi.org/10.1145/3295750.3298937>

1 INTRODUCTION

A Systematic Review, as a formal approach to document reviews, is a Recall oriented task [21], that appears in most disciplines. In their extreme forms, within evidence based medicine and legal e-discovery [30], *all* relevant documents *must* be found to be confident that decisions are being made in the light of all possible data, *and* that no data is missed. As an activity, a systematic review is usually performed by experts, and usually under very tightly controlled parameters that have been prescribed as the task was assigned. In practice, however, systematic reviews might be spread across multiple people as a collaborative search activity [18], and is typically performed across a complex multi-stage process [23]. Further, multiple people with different skills and expertise often take different roles at different stages.

Systematic reviews must be rigorously performed and are currently laborious and repetitive; they must be both sufficiently inclusive and comprehensive in order to include all related work. Further, researchers must then find, comprehend, extract, and integrate data from within these results. Our research questions were:

RQ1 What is the nature of the work task, and its sub-tasks?

RQ2 What opportunities exist to support the work task with search systems?

We aimed to reveal the full nature of the Systematic Reviews, as a Work Task [25], or indeed as a series of multiple Work Tasks. In contrast to published documentation on how systematic reviews *should* be performed, this paper’s main contribution lies in presenting a detailed Cognitive Work Analysis [40] of actual working practices around systematic reviews (Section 4), based on a focused-ethnography study [29]. Our results lay the foundations for future research into the design of search systems that support this high recall collaborative work task.

2 RELATED WORK

We begin by contextualising Systematic Reviews as a Work Task, where Work Tasks are typically defined as the larger task involving information use, that typically create the information need that leads to information seeking [25]. We then highlight the benefit of exploring a medical case study, before describing our work.

2.1 Systematic Reviews as a Work Task

Systematic reviews involve a well established process [9], applied in most research disciplines, including our own community (e.g., Kelly & Sugimoto’s systematic review of IIR Evaluation [27]). Brereton *et al.* [7] acknowledge 10 stages of a systematic review that span the planning, execution, and documentation of systematic reviews: 1) plan research questions, 2) specify review protocol, 3) validate review protocol, 4) identify relevant research, 5) select primary studies, 6) assess study quality, 7) extract required data, 8) synthesize data, 9) write review report, 10) validate report. In comparison to a literature review, a systematic review is designed to 1) parameterise a literature review space to define what will be included and excluded, 2) survey *all* the available research that meets those criteria, 3) synthesise the studies’ combined data (e.g., through meta-analysis), and 4) present quantifiable recommendations based on the synthesised data [19]. While a literature review might outwardly look for possible extant literature that relates by any one criteria, a systematic review looks inwardly to find all the literature that matches all the prespecified criteria, to the exclusion of results that only partially meet the criteria.

In applying these stages to software engineering literature, Brereton *et al.* highlight that poor quality abstracts and lack of infrastructure make such reviews difficult. Thus, they need to be adapted to suit different domains. Athukorala *et al.*, for example, found that literature searching was a highly collaborative experience for the computer scientists they studied [3]. Papaioannou *et al.* studied the different search tactics used by social scientists in systematic reviews, noting that beyond reference lists, checking and expert contacts were needed to reach rigorous standards [32]. More recently, Booth performed an in depth systematic review of methodologies used in *qualitative* systematic reviews [6], noting the data extraction of comparable specific detail as an open challenge.

The stages of the systematic review task lend themselves to different roles, similar to the Prospector and Miner proposed by Golovchinsky *et al.* [18], where one person’s role is to find sources of information, and another person’s role is to extract data from them. In 2005, Harris studied the crucial role that a medical research librarian plays in the process, in collaboration with researchers on a project [22]. Similarly, Beverley *et al.* studied 11 different roles that may be performed by an information specialist in healthcare literature reviews [5]. It could be argued that Systematic Reviews are made up of a series of Work Tasks, such as selecting primary studies, reviewing papers, etc.; however, as our results further highlight below, each of the stages are closely integrated and depend on shared document artefacts.

2.2 Case Study: Medical Systematic Reviews

While systematic reviews are recognised across disciplines, in medical research and practice, systematic reviews represent critical work tasks. For example, “evidence-based medicine” is the practice of ensuring that the therapies proposed by clinicians to their patients are those best supported by existing medical evidence [33], which is essential for preventing unnecessary harm to the patients due to unsafe or inefficient methods [8]. Thus medical practitioners must constantly study medical literature in order to continually

update and revise their practice in the light of new evidence. Often there is a delay between new research being published and clinicians updating their methods, leading to a so-called “evidence-practice gap” [20]. Narrowing this gap is a significant challenge for medical practitioners. Indeed, ‘*It is unlikely that all [medical practitioners] will have the time, skills and resources to find, appraise, and interpret this evidence and to incorporate it into healthcare decisions.*’ [9, Sec. 1.2.1]. Systematic reviews are a key approach to overcoming this problem. By providing practitioners with summarised data, systematic reviews narrow the evidence-practice gap by removing the need for individual clinicians to do their own literature reviews; they can instead refer to the systematic review performed by other experts.

2.3 Systematic Review Tools

An important challenge and a limitation of systematic reviews is their currency—i.e., the extent to which they reflect the most up-to-date research. The highly rigorous nature of systematic review production makes the review-writing process a very time-consuming one. There can be a gap of anywhere between 2.5 and 6.5 years between new research being published and that research being incorporated into a review [26], and there is also commonly at least a year between completing the literature search and publishing the final review [12], meaning that systematic reviews are often out of date from the moment they are written, or missing key evidence.

Not surprisingly then, much research (including the background motivation of our own work) is focused on identifying opportunities for tool provision for this work task [3]. Fabbri *et al.* [14], for example, have produced a tool for text mining content from found relevant papers, and visualising the results for the systematic reviewer. It is hoped that replacing human effort with machine effort during systematic review production will vastly reduce this latency, improving the relevance of the systematic reviews thus produced [12, 38]. Across the tasks, the challenge lies in identifying ways of automating elements of the review writing process while still maintaining the high standards of impartiality required in a systematic review [4, 39]. Currently, this impartiality is ensured by never relying solely on automated systems for any part of the review production process. Most stages of the process involve two or more workers operating in tandem, and it is recommended that replacing a *single* worker with an automated system should still ensure unbiased outputs [31]. Current systematic review support tools are largely designed, therefore, as decision support tools rather than as tools to replace humans altogether (e.g., [16, 28]).

3 FOCUSED-ETHNOGRAPHY STUDY

In this section, we describe our method for investigating systematic reviews, and introduce the context of our case study.

3.1 Methodology and Setting

A three-month ethnographic study was designed to understand the work of a medical research group involved in producing systematic reviews. While ethnographic studies may take years to conduct, studies of shorter duration can be sufficient to understand the environment, systems and practices and to provide the foundations for more in-depth studies [10]. Such shorter studies are particularly

appropriate for *design ethnography*, the application of ethnographic methods to systems design. They enable designers to account for the *social characteristics* of the system being studied, rather than focusing solely on the functional characteristics [11]. The aim of our study was to provide as complete picture of the group’s work as possible, revealing both the movement of artefacts between participants and the tight interdependence of sub-tasks.

Because systematic reviews are particularly important in medical research and practices (Section 2) we sought to engage a research group that is actively contributing to the Cochrane Collaboration.¹ The Cochrane Collaboration is an international organisation founded in 1993 to establish standards for the production of medical systematic reviews and make them widely available [9].

We approached the international Cochrane Schizophrenia Group (CSzG)² and conducted the study at their headquarters in the Institute of Mental Health at the University of Nottingham. Most systematic reviews produced by the group are written by volunteers who engage on a temporary basis. The core group retains a small number of editors who coordinate the writing process and select the topics for systematic reviews. The group’s headquarters occupies part of a shared office in the Institute of Mental Health, alongside a number of other mental health-related research groups, that provides a working space for local volunteer reviewers. These individuals and their roles are described in detail below.

3.2 Data Collection and Analysis

One of the authors spent three months working full-time at the group’s office, integrated with the review team. The author assisted with the data extraction for systematic reviews, while observing and interviewing selected members of the group. The data was collected through informal discussions and *ad hoc* interviews throughout January to March 2018. In particular, we followed four participants who play key roles in the group: The group’s Coordinating Editor (P1) and Information Specialist (P2), the director of a consultancy firm Review Solutions (P3), and a Medical Student who was engaged in producing a systematic review for their dissertation (P4). The gathered data included field observation notes, interview recordings and transcripts, screenshots of computer-based work, and examples of physical artefacts produced during the group’s work.

In order to gain an in-depth understanding of the *process* and the *roles*, we analysed the data using Cognitive Work Analysis (CWA) [40]. Where other analysis methods are either *descriptive* (presenting what *is* done) or *prescriptive* (describing what *should* be done), CWA represents a *formative* approach and describes the system in terms of the constraints on action it imposes (i.e., it shows what *may* be done). Thus the method is well aligned with our intent to understand the requirements for the enhancement of the computerised support.

CWA is made up of five sub-analyses (termed *phases*), that are conducted in a recommended order to capture multiple facets of work: 1) Work Domain Analysis, 2) Control Task Analysis, 3) Strategies Analysis, 4) Social Organisation and Cooperation Analysis, and 5) Worker Competencies Analysis. Each phase builds on the results of the previous phases to construct a complete picture of the system

under analysis. Analysis was performed in the first instance by the lead author, whose findings were then checked and discussed by the others.

A recent extension of CWA proposes modifying the phases for team-based Computer-Supported Collaborative Work [1]. However, that technique is more suited for same-time same-place collaboration [2]. We discovered early in our study that the team’s multi-stage process was set up and managed so that the work is more *social* [13] than collaborative [18], and so we proceeded with the original CWA approach [40].

4 FINDINGS

In this section we present the findings of the ethnographic study through the lens of the five CWA phases. We group our findings thematically according to the key work tasks we uncovered, rather than proceeding strictly according to the CWA order. Within discussions of specific work tasks we follow the relevant CWA phases in sequence. In our discussion we adopt the following terminology: a *report* refers to an individual document presenting data of interest (e.g., a journal article); a *study* is a collection of all reports related to the same underlying clinical trial; and a *review* designates a systematic review summarising several studies.

4.1 Nature of the Group’s Work

We begin with a broad description of the group’s work based on the first two phases of CWA and then define the Work Tasks and analyze them in depth.

4.1.1 Work Domain. The first phase of CWA is the Work Domain Analysis (WDA) that focuses on the physical and environmental constraints of the system used by CSzG. This analysis is important because a system’s function is constrained not just by the tasks that must be performed, but also by the environment they must be performed in. The results of the WDA are summarized in Table 1. In general, the CSzG group produces and publishes medical systematic reviews relating to the treatment of schizophrenia. Their work involves a mix of physical and digital artefacts. They use face-to-face and virtual communication, both constrained by their shared office space. The workers make use of physical artefacts and face-to-face communications wherever possible, but certain interactions are required to take place via computer.

4.1.2 Control Tasks. During the Control Task Analysis (ConTA) we identified distinct Work Tasks that are performed to complete the systematic reviews. We organise our findings based on prominent aspects of the Work Tasks that are of interest to our research. The summary of identified Work Tasks is presented in Table 2. ConTA aims to describe all the tasks performed within a system, taking no considerations of *how* they are done or by *whom*. In the later stages of our analysis it became apparent that, in practice, several of the lower level tasks are combined to form a few larger tasks. We broadly group these individual tasks under two processes: a “Search and De-duplication Process”, and a “Data Extraction Process”. Some tasks occur in both processes while others are specific to only one.

In brief, the Search and De-duplication process involves the group’s information specialist, who maintains and constantly updates the group’s database of schizophrenia-related studies. The

¹<https://cochrane.org>

²<https://schizophrenia.cochrane.org>

Table 1: Work Domain Analysis of CSzG

Abstraction level	Domain description
Functional purposes <i>Objectives & external limits</i>	Review production; Review dissemination; Use time & money efficiently; Cochrane Handbook
Abstract functions <i>Fundamental laws, principles, values</i>	Impartiality; Transparency; Rigour; Comprehensiveness; Uniqueness; Work-hours per review; Qualification of reviewers
Generalised functions <i>Functions enabling fulfilment of purposes</i>	Intra-team communication; Recording past reviews; Peer review; Screening; Data extraction; Meta-analysis; Study de-duplication; Translation of non-English documents
Physical functions <i>Capabilities & limits of physical objects</i>	Word processing; Highlighting of text; Annotation of text; Search for studies/reports; Record data; Question asking; Computer networking
Physical forms <i>Physical artefacts and tools</i>	PDF documents; Printed documents; Data extraction forms; Pens & highlighters; Post-it (sticky) notes; Skype; WeChat; MS Word; MS Excel; MS Access; MeerKat; Review Manager; RevMan HAL; Information specialist; Editors; Reviewers; Review Solutions consultancy; Chinese government firewall

Table 2: Key Work Tasks in CSzG

Work Task	Description
Register update and maintenance	Updating the study-based register; Refining the ontology used to code the register's contents.
Reviewer training	Providing training, as a service, for new reviewers, teaching skills needed to conduct a systematic review.
Study de-duplication	Determining which individual research reports describe the same underlying study in the database.
Screening	Determining the relevance of provided studies.
Data extraction	Obtaining required qualitative and quantitative data from relevant studies.
Review compilation	Entering results of data extraction into RevMan; Writing conclusions.

Data Extraction process takes place during the production of an actual review, and involves extracting data from medical research reports provided by the information specialist and recording them.

4.1.3 Cooperative, but individual. Before proceeding to describe these two processes, we first give a general account of the social and organisational characteristics of the group as a whole. We identified three broad classes of worker within the group: *Editors, Authors,*

and *Trials Search Coordinators*. Referring to the list of participants earlier, P1 is an Editor, P2 is a Trials Search Coordinator, and both P3 and P4 would be considered Authors (as would our researcher).

A key research question for us was about the possible *cooperative* nature of systematic review production. While the work processes of systematic reviews involve multiple individuals, it is not clear whether and how the cooperation manifests itself in practice. Social Organisation and Cooperation Analysis of CSzG's work indicated that the group's work does not fit conventional notions of collaborative or cooperative work. There were no organised teams of individuals conducting exactly the same tasks at the same time. Data extraction, for example, seems a suitable candidate for cooperative work. Instead, by the formal protocols, it was required to be performed individually in order to avoid bias. This makes the CSzG work more *coordinated* than cooperative or collaborative [34]. That said, Authors would often confer with others outside of their own review, in order to get or provide help and advice, particularly their associated Editors. In this sense, the work of producing a systematic review is more social [13] than collaborative, especially for specific tasks. In principle, several of the tasks could benefit from the development of systems to improve collaboration, but they would have to be carefully assessed to ensure that no biases are introduced into the end product. This finding highlights the importance of access permissions in collaborative search [18].

Cooperation was, however, observed in two cases: processing reports written in Mandarin and quality assurance at the end of the extraction process. In the first case the group had to work with reports published in Mandarin. As there are many studies in Schizophrenia research published in non-English languages, for an English-speaking research group the data in them are '*locked behind a language barrier*' (P1). For that reason, CSzG works with a Chinese consultancy firm to extract data from Mandarin-language reports and provide them to reviewers. That requires close communication between the study author, the consultancy's director based in the CSzG office (P3), and the consultancy staff assigned to perform the data extraction. This communication is complicated by the fact that all the RS staff except for the RS Director are based in China and work remotely. The second case of cooperation was observed at the end of the data extraction task. In principle, this task is conducted by two reviewers entirely separately, who then come together to discuss the extracted data, check one another's results, and agree on the final data to be included in the review.

4.2 Search and De-duplication Process

The first stages of a systematic review involve conducting the literature search, screening the results to filter out non-relevant reports, and grouping relevant reports that relate to the same clinical trial. In CSzG these tasks are conducted by the group's information specialist (P2), as part of the Search and De-duplication Process. We note that this set up represents a significant departure from the traditional systematic review methodology. We defer providing more details and commenting on the implications until the discussion.

4.2.1 Strategies for Search and De-duplication. The Strategies Analysis phase of CWA is concerned with describing the different possible approaches to completing the control tasks identified during ConTA. The Search and De-duplication Process, in particular,

involves three control tasks: Register Update and Maintenance, Screening, and Study De-duplication.

The CSzG strategy for Register Update and Maintenance is thoroughly described in [35], as the group has pioneered this methodology. In CSzG, the group’s information specialist (P2) runs a monthly literature search using a previously-approved search protocol, performs initial screening of the results to ensure they relate to schizophrenia, and records them in a database along with their broad PICO³ characteristics and their full text in PDF form:

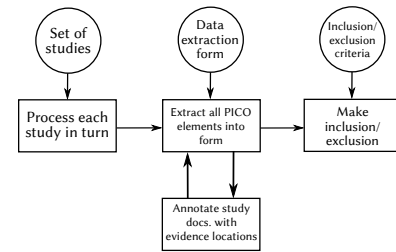
So, to have this register updated, we’ve used 70 different databases, like your medical databases. [...] We have MEDLINE, Embase. We do the searches in 10 databases every month. It’s automatically done, the majority of it, and I receive the records. I screen them to see if they are randomised, if they [say] schizophrenia. If so, I am adding to our database and I extract those metadata, that I told you. Which is participants, are they schizophrenia? Age group, and do they have a special problem, like depression and schizophrenia? Then, what are the interventions and what are the outcomes? (P2)

The search strategy employed by the group is fixed, in compliance with the requirements for rigorous and reproducible search results imposed by the Cochrane Handbook (e.g., [9, Sec. 6.1.1.2]). These regulations prevent the group from utilising more interactive or exploratory search strategies. In principle, the regular literature search strategy could be enhanced, but any changes to the methodology would have to be carefully studied to ensure that the required degree of transparency and reproducibility is retained.

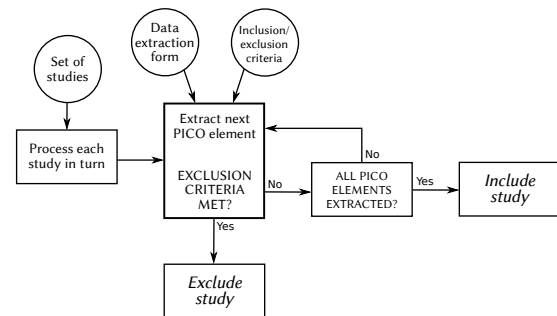
Screening is the task of filtering out results of the literature search that are non-relevant. Diagrams illustrating strategies for screening are shown in Figure 1. The general method for screening a particular report involves comparing the PICO characteristics of that report to the PICO characteristics of the systematic review. In the case of the current process, this is performed more generally, as described in the quote above: any report of a randomised controlled trial relating to schizophrenia is deemed relevant at this stage and included in the database. Reports are labelled in the database according to an informal taxonomy defined by the group itself, which consists of standardised names and spellings of drugs, outcome measures, and other data of interest. This enhances later retrieval of documents.

Finally, the de-duplication task is that of identifying which individual reports returned in the search results relate to the same underlying clinical trial. This is important, because often several publications will be made from the same original study, but, ‘Duplicate publication can introduce substantial biases if studies are inadvertently included more than once in a meta-analysis.’ [9, Sec. 7.2.2] Only one feasible strategy for study de-duplication was identified, shown in Figure 2.

4.2.2 Social Organisation and Worker Competencies. The fourth and fifth phases of CWA describe how workers are organised within the system being analysed, and what skills and abilities those workers require to perform their tasks. Conventionally in a medical systematic review, the responsibility for the described control tasks would be the responsibility of an Author, with the search task supported by a Trials Search Coordinator [9, Sec. 6.1.1.1]. In the CSzG,



(a) Naive approach



(b) Iterative approach

Figure 1: Strategies Analysis for Screening

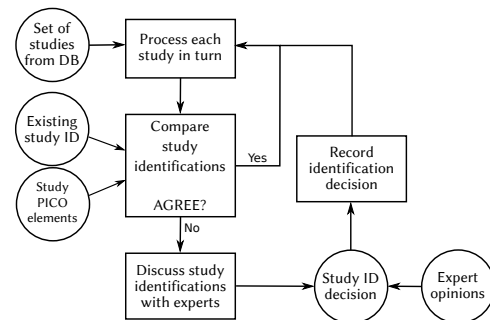


Figure 2: Strategies Analysis for Study De-duplication

however, these tasks are allocated to a single person, the information specialist. As will be described below, Authors still engage with some parts of the Search and De-duplication Process, but only at a later stage, and only in a confirmatory capacity.

Worker Competency Analysis typically breaks down the individual behaviours (required to perform a control task) into three categories: *skill-based* behaviours which are at the fundamental level and considered almost instinctive; *rule-based* behaviours which involve execution of learned sets of rules; and *knowledge-based* behaviours which require critical thinking and debate on the part of the actor in order to undertake the execution. Among the identified tasks, by far most complex is de-duplication, requiring a high degree of knowledge-based behaviours. Screening is rule-based, as the actor simply compares extracted PICO characteristics against those of the study.

³An acronym referring to the four major categories of qualitative data extracted from a study: Patient, Intervention, Control, Outcome.

4.3 Data Extraction Process

The Data Extraction Process seemed particularly demanding to those involved in the corresponding tasks. The Coordinating Editor (P1) described systematic reviewers as being ‘enslaved to the trapped data’, with reviewers ‘chiselling the mine’ to get at the data they needed. Both qualitative and quantitative data are extracted from the studies under review. The qualitative data consist of PICO characteristics describing the study in terms of the drugs that were involved, the number of patients studied, and similar, while the quantitative data include the actual results of the study in terms of measured changes in outcomes. As shown in Table 2, this process involved the control tasks of Screening, Study De-duplication, Data Extraction, and Review Compilation. This process is almost exclusively performed by Authors, with advice from Editors or a Trials Search Coordinator when requested.

4.3.1 Data Extraction Strategies. The first two control tasks are also carried out during Search and De-duplication as described in the previous section. The strategies remain the same, but the goals are slightly different. Study de-duplication is only conducted to verify the decision of the information specialist to include the study in the review, rather than to conduct the search from scratch. De-duplication decisions are in fact constantly up for debate, and may be changed at any time with a sufficient justification:

[...] whatever conclusion [we reach], we [re]apply on our register. So, whatever exists in our register is the accumulation of all of the efforts that have been done by reviewers, editors, me, people who have realised there is an error and report it. Whoever reports any error, I just correct it here. (P2)

Screening, on the other hand, is more precise here than during Search and De-duplication. In this process, the screening involves consideration of more detailed characteristics and inclusion criteria. The author embedded in the group participated in a systematic review where one study was excluded because its outcomes were not specific enough to be clearly related to the purpose of the review:

So, if you look in diagnosis, it doesn't really specify if the patients had aggression. It was just an exacerbation of schizophrenia. So we were like, "You can't say psychosis induced aggression." (P4)

It is the screening during the data extraction process that revealed the naive screening strategy shown in Figure 1a. This approach was adopted by a first-time review author with minimal training in systematic review production. Commenting on one data extraction form, she said,

[P1] said, although I've done this one a bit more fully, he said usually what happens is, with the excluded ones, they fill in the things until it gets excluded, and then they stop filling it in [...] So, technically I could have stopped there and it would have been fine. (P4)

The author's inexperience led her to obtain the vast majority of the study's characteristics and only then make a decision to exclude it from the review.

Data Extraction is a new control task, only performed during this process. As the name suggests, it is this task that requires the majority of the effort during the Data Extraction Process. The task consists of identifying all of the qualitative and quantitative data contained in the studies processed for the systematic review, and recording them. Strategies for this task are shown in Figure 3. In

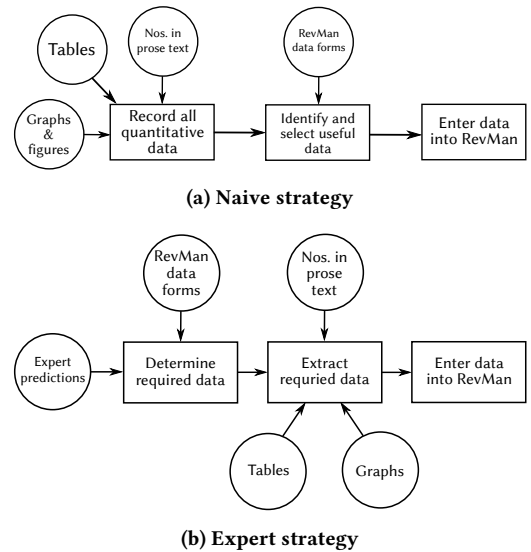


Figure 3: Strategies Analysis for Data Extraction

the group, the preferred practice is to print physical copies of the reports in a study, along with a data extraction form, and to use these as reference materials for data extraction rather than working with digital media.

In all cases, data are first extracted into a specially-designed data extraction form (see Figure 4 in the appendix for an example), and once the form is completed, the data are then entered into the review writing software. The data extraction form specifies precisely which qualitative data are required for the review, and the data extraction task consists simply of filling in this form with evidence drawn from the study. Quantitative data extraction is more complicated that this, however. Only certain data are actually required for the review, to support the meta-analysis, but an inexperienced reviewer will not know in advance which data these are. Referring back to the review author, she experienced exactly this problem:

P4 *[Entering the extracted data] was the longest part, because so much had to get deleted, ultimately [...] So, a lot of the stuff that [report authors] put isn't needed [for the review], and so they put in baseline figures and we don't use those. Skewed data goes.*

Int. *Okay, so you've pulled out all the numbers?*

P4 *Yes, you've pulled out absolutely everything.*

Int. *Then when you go to enter it, you find that you don't actually need all of the numbers.*

P4 *Yes, it takes a lot of careful checking.*

The data are simply extracted onto sheets of paper (Figure 5) and then entered later into the review writing software. In principle it should be possible for a more experienced review author to predict which quantitative data are going to be needed for a particular review and only extract those, hence the alternative strategy shown in Figure 3b.

A vital part of all strategies for data extraction is the annotation of the source documents to indicate the location of the evidence

for the data in the forms. This annotation may take the form of highlighting sentences or phrases (see Figure 6), or by placing small numbered marks in the forms that are then referred back to. These annotations serve to make it much easier for other reviewers to verify that the extracted data are correct.

Yes, so, [P1] makes [the data extraction form], he had to be like, 'Yes, highlight where you've got things from, annotate where you've got the things from.' [...] [In] a lot of them I put asterisks, numbers, just to know that I can go back and see exactly where I've got the stuff from. (P4)

It should be noted that this reliance on using physical artefacts for data extraction may not be a widely adopted practice across systematic reviewers. There are different software solutions available to support systematic reviewers in extracting and storing data,⁴ but flexibility in the process and the data means that the group prefers to work on paper. The variability of data, types of data, and formats of presentation mean that paper adds flexibility in a way that is often complex to record in digital forms. The most obvious example of flexibility requirements is collaboration with Review Solutions to extract data from Mandarin-language studies prevents the use of many of these tools. The director of Review Solutions commented,

The reason we haven't used [Covidence] is because my team is mostly based in China, and the government has a firewall, which makes [...] Covidence incredibly slow. [...] There's another tool—Apparently, it's more powerful than Covidence; it's called Distiller [...] but then we had the same trouble. Sometimes we simply can't even get on to the website. (P3)

4.3.2 Social Organisation and Worker Competencies. The Social Organisation and Cooperation Analysis for this process uncovered a range expertise levels amongst review authors, from fully-qualified medical doctors, to medical students, to entirely non-medical workers such as the embedded researcher. In addition to these differing levels of domain expertise, authors also possessed varying levels of expertise in the systematic review process. As such, the authors had and required varying levels of training in data extraction.

It quickly became apparent that this variance in expertise was not an obstacle to the production of systematic reviews of an appropriate quality. The group's coordinating editor *prefers* to make use of less-qualified people to conduct reviews, citing the use of full time fully-paid medical doctors as 'a waste of NHS money'. Training is provided by the coordinating editor on an informal basis, but there is also a lot of "learning on the job":

P4 *At the beginning, I was really overwhelmed. I did not understand how I was going to be able to do this in such a short time period. [...] I don't really like to make a big thing of asking help all the time [...] [but] as soon as you start doing it, it gets easier and easier [...]*

Int. *Okay, so a lot of learning by example, would that be fair?*

P4 *Yes, and then also [P1] gave loads of advice, but we've also got [P3] here [...] she's done some stuff and then [P2] is really good at all the search stuff, and even then the people that come in and out. A lot of the time they'll literally just pop in, I'll ask them a question, and they'll disappear. I have no idea who they are, but they were very helpful.*

The coordinating editor described his own role in the process as to be 'breathing down the neck' of the reviewers to guide them in how to write their reviews to an appropriate standard. He also runs an annual training seminar on how to conduct systematic reviews, but it was not possible to observe this during the study.

Data extraction is conventionally organised to involve two separate review authors working in tandem. The recommended method is for each author to separately conduct data extraction for the studies in the review, and for them to then come together and compare results [9, Sec. 7.6.2]. In this way, the authors can check each others' work to improve accuracy of data extraction. In practice, this may not always be performed "by the book"; in the systematic review we participated in, data extraction had been completed by the first author (P4), who then gave their results to the embedded researcher and simply asked them to 'double check' the data.

In terms of worker competencies, de-duplication and screening remain much the same as they were described previously. Screening becomes more knowledge-based in this process, however, because the decisions to be made are more complex. Data extraction itself requires a wide mix of skill- and knowledge-based behaviours: certain PICO elements can be identified almost by "pattern matching" on the part of the review author, while other data require much interpretation of the source document before they can be successfully extracted, especially when it comes to extracting qualitative data.

5 DISCUSSION

This section draws together an initial set of conclusions from the CWA analysis, organised according to our research questions.

5.1 The Key Tasks (RQ1)

Broadly, we would typically classify the systematic review task as one involving Exploratory Search [41], well beyond the query-response paradigm. *Exploration*, however, is actively discouraged in favour of comprehensive review of results, by creating *processes* that make it impartial and procedural.

5.1.1 The Pre-Search Problem. The first stage of their processes involved the procedural retrieval and pre-categorisation of new literature into a purpose built taxonomy, which is performed entirely separately from an ongoing systematic review task. By creating subscriptions to digital libraries, and performing pre-determined manual searches, the team reviews all new publications for how they relevant to their expertise, and classifies the results for future reviews. This represents an information monitoring task, and is perhaps an ideal opportunity for more advanced Slow Search [36] systems. Further, this pre-search stage involves document inspection and judging relevance of a straight list of results, which is not unlike a standard, yet comprehensive, version of a very straight information retrieval task.

The larger *work task* of this phase is to classify of results into different parts of the taxonomy, designed to make subsequent search tasks more procedural. Indeed, the work is a lot like those performed to create a TREC test collection. However, in this case they are creating an intermediary search system, such that people performing systematic reviews further down the process do not actually explore the literature available in digital libraries, but retrieve pre-classified data from pre-defined queries.

⁴Covidence was mentioned by P3 [<https://www.covidence.org/home>]

5.1.2 The Search Problem. Once a systematic review begins, the work is still executed in separate phases, often by different people. One stage involves the more exploratory *searching*, and the other involves more investigative *inspection* of results.

The first phase involves the identification, only, of search queries for a systematic review. This activity focuses on identifying the queries that are relevant, in order to meet the requirements of the systematic review. This may involve interactive querying and making relevance judgements of result sets, but it rarely involves what we would often consider a core part of the information retrieval process: finding information.

The second phase of the systematic review involves almost no searching at all, and instead involves the comprehensive analysis of the results returned from the queries built (by another person) in the first phase. This person cannot, or should not, seek more results or explore related terminology, they must simply examine the results provided to them by the intermediary search system. In the examples from our observations, a study that included symptoms of acute exacerbation was not included, because it was not specifically controlling for *aggression* as a variable. Indeed, the reports produced as a result of the systematic review are expected to identify works that are excluded; these results are *found*, judged as *relevant*, but not *included* in the review. These findings help to elaborate on part of the *Collaborative Search role* that Golovchinsky *et al.* referred to as a Prospector [18].

5.1.3 The Post-Search Problem. The role of a Miner [18] in our case study, has to make very detailed relevance judgements based upon a key challenge: Data Extraction. Here, the quantitative data must be examined by the main systematic reviewer to determine precisely which reported outcomes are usable or not. Qualitative data extraction was observed to be generally an exercise in pattern matching, and most of the data extraction involved simply searching for key words or phrases in the text that indicated a sentence contained PICO data. For future systems, automated quantitative data extraction could prove to be a hard challenge, as the data needed for the review may only consist of one or two rows of a single table in an entire report, and at present the reviewers must do a lot of manual work to identify exactly what is needed.

The heterogeneous nature of the group's document collection also makes the work of data extraction harder; there were a variety of document types in the corpus, ranging from journal articles to conference abstracts to doctoral theses. There were also a variety of qualities of PDF documents, some of which have very unreliably transcribed text which will make automated extraction inconsistent and more prone to errors. These challenges are similar to those identified by Brereton *et al.* [7].

5.2 HCIR Opportunities (RQ2)

Our investigation identified a series of opportunities for HCIR systems to support systematic reviews through automation of tasks that were described as laborious and repetitive. One initial observation is that this multi-stage process may be facilitated better by systems that, overall, model search stages explicitly [23, 24]. The breadth of expertise observed in those conducting systematic reviews indicates that great care must be taken when designing any technologies for systematic review automation to enable as many

people as possible to participate in the review production process, particularly in data extraction. There is already interest in enabling less-experienced volunteers to participate in certain stages of review production (e.g., [37]), and so care should be taken when developing new systems to ensure they are usable by novice reviewers.

In the pre-search stage, systematic review process might be best supported by tools designed for taxonomy management, than for exploratory search. This role involves work more like a librarian, cataloguing data for their team, than a reviewer/searcher.

In the main systematic review stage, the work is more focused on result-dataset comprehension task, such that tools could better support a searcher in knowing a) what portion of the available results they have retrieved, and b) how they relate to parts of the taxonomy. One tool that made developments in this area was Querium [17], which made efforts to help searchers understand the relationships between each of their queries, returned results, and the full dataset.

The post-search data extraction problem is one that is more closely related to automatic summarisation. In practical terms, a tool for systematic review data extraction would need to be capable of coping with arbitrary document layouts and PDF qualities, for example. More importantly, however, the data extraction problem is a varied and interactive task involving retrieving data from within documents, rather than documents themselves. Given what was observed of people using paper artefacts, this task could benefit from e.g. interactive machine learning interfaces like Cueflik [15], that would allow searchers to dynamically specify the types of data they are intending to extract for the systematic review.

6 CONCLUSIONS

In this paper we have presented a Cognitive Work Analysis, based on a novel focused-ethnography study of Medical Systematic Reviewing, as a rich and well-structured case study of such a Work Task. In its premise, we might consider Systematic Reviews to involve Exploratory Search activities. In practice, however, we find that well structured systematic review processes can involve a series of constrained search tasks, performed by different people, none of which involves exploratory search as we would normally describe it. Systematic Reviews, in our case study, were facilitated by a pre-search task that categorised new documents asynchronously from any actual systematic review. Notably, all subsequent people in the process work with these pre-classified documents without broadly searching online digital libraries for new documents. Perhaps the most exploratory actions are taken by a different person who identifies and then dictates the queries that will be used by the main systematic reviewer, but does very little exploring of the actual results. This final stage is then largely a data extraction task rather than a search task, reviewing a potentially large linear list of results. We find that most of these processes are purposefully delineated such that collaboration is coordinated by process, rather than performed as a synchronised co-located activity. There are opportunities, however, to support each of the subtasks involved in systematic reviews with different types of search tools. Alternatively, a system designed to support individuals in achieving the entire process alone would likely benefit from facilitating the separation of processes observed in our case study.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/M000877/1]. We would like to thank all of our study participants for taking the time to talk to us and explain their work, and the Cochrane Schizophrenia Group for affording us valuable desk space. We are also very grateful to our anonymous reviewers for their feedback and suggestions.

Data access: Due to concerns over preserving the anonymity of our participants, supporting data cannot be made publicly available.

REFERENCES

- [1] Maryam Ashoori and Catherine Burns. 2013. Team Cognitive Work Analysis: Structure and Control Tasks. *Journal of Cognitive Engineering and Decision Making* 7, 2 (2013), 123–140. <https://doi.org/10.1177/155534312445577>
- [2] Maryam Ashoori, Catherine M. Burns, Barbara d'Entremont, and Kathryn Montahan. 2014. Using team cognitive work analysis to reveal healthcare team interactions in a birthing unit. *Ergonomics* 57, 7 (2014), 973–986.
- [3] Kumaripaba Athukorala, Eve Hoggan, Anu Lehtiö, Tuukka Ruotsalo, and Giulio Jacucci. 2013. Information-seeking Behaviors of Computer Scientists: Challenges for Electronic Literature Search Tools. In *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries (ASIST '13)*. American Society for Information Science, Silver Springs, MD, USA, Article 20, 11 pages. <http://dl.acm.org/citation.cfm?id=2655780.2655800>
- [4] Elaine Beller, Justin Clark, Guy Tsafnat, Clive Adams, Heinz Diehl, Hans Lund, Mourad Ouzzani, Kristina Thayer, James Thomas, Tari Turner, Jun Xia, Karen Robinson, Paul Glasziou, and founding members of the ICASR group. 2018. Making progress with the automation of systematic reviews: Principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). *Systematic Reviews* 7, 1 (2018), 77–84. <https://doi.org/10.1186/s13643-018-0740-7>
- [5] C.A. Beverley, A. Booth, and P.A. Bath. 2003. The role of the information specialist in the systematic review process: a health information case study. *Health Information & Libraries Journal* 20, 2 (2003), 65–74. <https://doi.org/10.1046/j.1471-1842.2003.00411.x>
- [6] Andrew Booth. 2016. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. *Systematic Reviews* 5, 1 (04 May 2016), 74. <https://doi.org/10.1186/s13643-016-0249-x>
- [7] Pearl Brereton, Barbara A. Kitchenham, David Budgen, Mark Turner, and Mohamed Khalil. 2007. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software* 80, 4 (2007), 571 – 583. <https://doi.org/10.1016/j.jss.2006.07.009> Software Performance.
- [8] Archie Cochrane. 1972. *Effectiveness and Efficiency: Random Reflections on Health Services*. The Nuffield Provincial Hospitals Trust.
- [9] Cochrane Collaboration. 2011. *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0*. <http://handbook-5-1.cochrane.org>
- [10] Andy Crabtree. 2003. *Designing Collaborative Systems: A Practical Guide to Ethnography*. Springer-Verlag, London.
- [11] Andrew Crabtree, Mark Rouncefield, and Peter Tolmie. 2012. *Doing Design Ethnography*. Springer-Verlag, London.
- [12] Julian H. Elliott, Tari Turner, Ornella Clavisi, James Thomas, Julian P.T. Higgins, Chris Mavergames, and Russell L Gruen. 2014. Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap. *PLoS Medicine* 11, 2 (2014), 1–6. <https://doi.org/10.1371/journal.pmed.1001603>
- [13] Brynn M. Evans and Ed H. Chi. 2010. An elaborated model of social search. *Information Processing & Management* 46, 6 (2010), 656–678.
- [14] Sandra Fabbri, Elis Hernandez, Andre Di Thommaso, Anderson Belgamo, Augusto Zamboni, and Cleiton Silva. 2013. Using Information Visualization and Text Mining to Facilitate the Conduction of Systematic Literature Reviews. In *Enterprise Information Systems, José Cordeiro, Leszek A. Maciaszek, and Joaquim Filipe (Eds.)*. Springer Berlin Heidelberg, Berlin, Heidelberg, 243–256.
- [15] James Fogarty, Desney Tan, Ashish Kapoor, and Simon Winder. 2008. CueFlik: interactive concept learning in image search. In *Proceedings of the sigchi conference on human factors in computing systems*. ACM, 29–38.
- [16] Allison Gates, Cydney Johnson, and Lisa Hartling. 2018. Technology-assisted title and abstract screening for systematic reviews: A retrospective evaluation of the Abstrackr machine learning tool. *Systematic Reviews* 7 (2018), 45. <https://doi.org/10.1186/s13643-018-0707-8>
- [17] Gene Golovchinsky, Abdigani Diriye, and Tony Dunnigan. 2012. The future is in the past: designing for exploratory search. In *Proceedings of the 4th Information Interaction in Context Symposium*. ACM, 52–61.
- [18] Gene Golovchinsky, Pernilla Qvarfordt, and Jeremy Pickens. 2008. Collaborative information seeking. *Information Seeking Support Systems* (2008), 47–48.
- [19] David Gough, Sandy Oliver, and James Thomas. 2017. Introducing Systematic Reviews. In *An Introduction to Systematic Reviews* (2nd ed.), David Gough, Sandy Oliver, and James Thomas (Eds.). Sage, London, 1–18.
- [20] Jeremy M. Grimshaw, Martin P. Eccles, John N. Lavis, Sophie J. Hill, and Janet E. Squires. 2012. Knowledge translation of research findings. *Implementation Science* 7, 1 (2012), 50–67. <https://doi.org/10.1186/1748-5908-7-50>
- [21] Maura R Grossman, Gordon V Cormack, and Adam Roegiest. 2016. TREC 2016 Total Recall Track Overview.. In *TREC*.
- [22] Martha R. Harris. 2005. The librarian's roles in the systematic review process: a case study. *Journal of the Medical Library Association* 93, 1 (2005), 81.
- [23] Hugo C Huurdeman and Jaap Kamps. 2014. From multistage information-seeking models to multistage search systems. In *Proceedings of the 5th Information Interaction in Context Symposium*. ACM, 145–154.
- [24] Hugo C Huurdeman, Max L Wilson, and Jaap Kamps. 2016. Active and passive utility of search interface features in different information seeking task stages. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*. ACM, 3–12.
- [25] Kalervo Järvelin and Peter Ingwersen. 2004. Information Seeking Research Needs Extension towards Tasks and Technology. *Information Research: An International Electronic Journal* 10, 1 (2004), n1.
- [26] Siddhartha R. Jonnalagadda, Pawan Goyal, and Mark D. Huffman. 2015. Automating data extraction in systematic reviews: a systematic review. *Systematic Reviews* 4, 1 (2015), 78. <https://doi.org/10.1186/s13643-015-0066-7>
- [27] Diane Kelly and Cassidy R. Sugimoto. 2013. A systematic review of interactive information retrieval evaluation studies, 1967–2006. *Journal of the American Society for Information Science and Technology* 64, 4 (2013), 745–770. <https://doi.org/10.1002/asi.22799>
- [28] Svetlana Kiritchenko, Berry De Bruijn, Simona Carini, Joel Martin, and Ida Sim. 2010. ExaCT: Automatic extraction of clinical trial characteristics from journal publications. *BMC Medical Informatics and Decision Making* 10 (2010), 56. <https://doi.org/10.1186/1472-6947-10-56>
- [29] Hubert Knoblauch. 2005. Focused Ethnography. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 6, 3 (2005). <https://doi.org/10.17169/fqs-6.3.20>
- [30] Douglas W. Oard, Jason R. Baron, Bruce Hedin, David D. Lewis, and Stephen Tomlinson. 2010. Evaluation of information retrieval for E-discovery. *Artificial Intelligence and Law* 18, 4 (01 Dec 2010), 347–386. <https://doi.org/10.1007/s10506-010-9093-9>
- [31] Alison O'Mara-Eves, James Thomas, John McNaught, Makoto Miwa, and Sophia Ananiadou. 2015. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews* 4, 1 (2015), 5–27. <https://doi.org/10.1186/2046-4053-4-5>
- [32] Diana Papaioannou, Anthea Sutton, Christopher Carroll, Andrew Booth, and Ruth Wong. 2010. Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information & Libraries Journal* 27, 2 (2010), 114–122. <https://doi.org/10.1111/j.1471-1842.2009.00863.x>
- [33] David L. Sackett. 1997. Evidence-based medicine. *Seminars in Perinatology* 21, 1 (1997), 3–5. [https://doi.org/10.1016/S0146-0005\(97\)80013-4](https://doi.org/10.1016/S0146-0005(97)80013-4)
- [34] Chirag Shah. 2009. Toward collaborative information seeking (CIS). (2009). arXiv:0908.0709
- [35] Farhad Shokraneh and Clive Elliott Adams. 2017. Study-based registers of randomized controlled trials: Starting a systematic review with data extraction or meta-analysis. *BiolImpacts* 7, 4 (2017), 209–217. <https://doi.org/10.15171/bi.2017.25>
- [36] Jaime Teevan, Kevyn Collins-Thompson, Ryan W White, Susan T Dumais, and Yubin Kim. 2013. Slow search: Information retrieval without time constraints. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*. ACM, 1.
- [37] James Thomas, Anna Noel-Storr, Ian Marshall, Byron Wallace, Steven McDonald, Chris Mavergames, Paul Glasziou, Ian Shemilt, Annelise Synnot, Tari Turner, and Julian Elliott. 2017. Living systematic reviews: 2. Combining human and machine effort. *Journal of Clinical Epidemiology* 91 (2017), 31–37. <https://doi.org/10.1016/j.jclinepi.2017.08.011>
- [38] Guy Tsafnat, Paul Glasziou, Miew Keen Choong, Adam Dunn, Filippo Galgani, and Enrico Coiera. 2014. Systematic review automation technologies. *Systematic Reviews* 3, 1 (2014), 74–89. <https://doi.org/10.1186/2046-4053-3-74>
- [39] Alexander Tsertsvadze, Yen Fu Chen, David Moher, Paul Sutcliffe, and Noel McCarthy. 2015. How to conduct systematic reviews more expeditiously? *Systematic Reviews* 4, 1 (2015), 160–166. <https://doi.org/10.1186/s13643-015-0147-7>
- [40] Kim J. Vicente. 1999. *Cognitive Work Analysis: Towards safe, productive, and healthy computer-based work*. Lawrence Erlbaum Associates, Mahwah, NJ.
- [41] Ryan W White and Resa A Roth. 2009. Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services* 1, 1 (2009), 1–98.

6

	Risperidone N=153	Quetiapine N=156	Placebo N=73
AIMS total score			
Mean baseline	1.1 ± 2.97	1.1 ± 2.70	0.8 ± 2.49
mean change from baseline @ endpoint	0.3 ± 0.2	-0.1 ± 0.2	-0.1 ± 0.3
SAS total score			
Mean baseline	0.9 ± 2.5	0.7 ± 2.0	0.7 ± 2.1
mean change from baseline @ endpoint	0.8 ± 0.2	-0.1 ± 0.2	-0.1 ± 0.3
BAS - basal severity of abnormality score from baseline @ endpoint [%]			
worsened	22 (15%)	10 (7%)	5 (8%)
unchanged	114 (78%)	115 (79%)	51 (77%)
improved	10 (7%)	20 (14%)	10 (15%)
Statistical comparison			
vs placebo	p=0.077	p=0.779	
vs quetiapine	p=0.007		

7

	Risperidone N=153	Quetiapine N=156	placebo N=73
No subjects with cut-point score > CAE	100 (65%)	97 (62%)	44 (60%)
Insomnia	29 (19%)	22 (14%)	17 (23%)
Headache	22 (14%)	18 (12%)	10 (14%)
Sedation	10 (7%)	15 (10%)	5 (7%)
Somnolence	4 (3%)	16 (10%)	2 (3%)
Dizziness	9 (6%)	16 (10%)	3 (4%)
Constipation	11 (7%)	8 (5%)	1 (1%)
Apathy	11 (7%)	1 (1%)	1 (1%)
Constipation	8 (5%)	14 (9%)	2 (3%)

Figure 5: Example quantitative data extraction

A EXAMPLES OF PHYSICAL ARTEFACTS

In this appendix we show some examples of the physical artefacts referred to in Section 4.

Trial ID	Kalati 2008	
Action	EXCLUDE → doesn't specify PR or aggressive pre-intervention	
Methods	Allocation: randomized Blindness: double-blind Duration: 6 week	
Participants	Funding: sponsored by AstraZeneca Diagnosis: DSM-IV diagnosis of schizophrenia with acute exacerbation N= 588 Age: 18-65 yrs Sex: M/F	
History:	Included: PANSS ≥ 70, PANSS ≥ 4 for at least either: delusions/conceptual disorganization/hallucinations/behavior suspiciousness/persecution CGI-S ≥ 4 Excluded: hospitalization for schizophrenia > 1 month immediately before randomization, known risk of suicide in last 12 months, previously enrolled in any study 4 weeks prior	
Interventions	1. Once daily quetiapine XR 400mg/day, 600mg/day or 800mg/day OR 2. twice daily quetiapine (R) 400mg/day OR 3. placebo	
Outcomes - able to use - list what was measured and how it was measured	Short term change in PANSS total score @ week 6 Efficacy effects on PANSS aggression/hostility cluster scores PANSS-EC	
Outcomes - unable to use - and WHY	efficacy on PANSS aggression/hostility cluster scores PANSS-EC	
Notes		

Figure 4: Example data extraction form

256 S.G. Potkin et al. / Schizophrenia Research 85 (2006) 254-261

The trial was conducted in accordance with current International Conference on Harmonization—Good Clinical Practice guidelines and the Declaration of Helsinki and its subsequent revisions. All patients deemed competent by the investigator provided written informed consent prior to study participation. If a patient was deemed not legally competent, then assent was obtained from the patient, and an authorized representative had to consent.

2.2. Patients

Patients were 18-64 years of age, with a DSM-IV diagnosis of schizophrenia (paranoid, disorganized, or undifferentiated type) or schizophreniform disorder (as defined by the Mini-International Neuropsychiatric Interview—Plus Version (MINI-Plus) (Sheehan et al., 1998). Enrolled patients were experiencing an acute exacerbation of their illness of recent onset (within 4 weeks) with prominent troublesome symptoms requiring hospitalization. Patients were required to have a score of ≥ 2 (at least two) of the following items on the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987): Hostility (P7), Excitement (P4), Tension (G4), Disorientation (G8), and Poor Impulse Control (G14), and a total score on these items of ≥ 47. In addition, a rating on the Clinical Global Impression of Severity (CGI-S; Guy, 1976a) of ≥ 5 (markedly ill) was required.

Exclusion criteria included any Axis I diagnosis, except abuse/dependence disorders; an Axis II diagnosis of mental retardation or borderline personality disorder; treatment-resistant schizophrenia; imminent risk for self-harm having received a depot antipsychotic within one dosing cycle prior to baseline; having received risperidone or quetiapine within 7 days prior to baseline; known allergy or sensitivity to either drug; evidence of a clinically significant or unstable disease, including a thyroid disorder not stabilized for at least 3 months.

2.3. Study medication/dosing

Patients received visually matching, over-encapsulated, comparative dissolution-tested study medication. Dosing regimens for risperidone and quetiapine were in accordance with the prescribing information for each drug and also reflected clinical and research practices for treating patients with acute exacerbations of schizophrenia (Lieberman et al., 2005; Stroup et al., 2003). From Day 1 through Day 5, study medication was administered according to a fixed schedule, with doses based on body weight. Patients randomized to risperidone were titrated from 1 mg/day to a target dose of 4 mg/day (≤ 70 kg) or 6 mg/day (> 70 kg) by day 5. Patients randomized to quetiapine were titrated from 50 mg/day to a target dose of 400 mg/day (≤ 70 kg) or 600 mg/day (> 70 kg) by day 5. Following Day 5, patients were maintained on the same dose of study medication for the remainder of the study except that investigators were able to increase the dose of quetiapine in a blinded fashion to 600 mg/day (≤ 70 kg) or 800 mg/day (> 70 kg) on day 8.

2.4. Concomitant medication

The use of other psychotropic medications was prohibited during the monotherapy phase of the study (Days 1-14). However, during this phase, short-acting, non-benzodiazepine hypnotics (e.g., zolpidem, zalcipon, zopiclone) for treating insomnia, and injectable lorazepam, sodium amytal, or midazolam for treating agitation or restlessness were permitted as needed. After Day 14, the investigator could prescribe any psychotropic medication deemed necessary, except specifically prohibited medications (drugs known to interact with the cytochrome P450 isoenzymes CYP2D6 and CYP3A4, and drugs with potential for benzodiazepine mesylate or meprobamate).

2.5.1. Efficacy

Primary efficacy was assessed using the Risperidone Questionnaire (RDQ) (Potkin et al., 2005). The RDQ rating yielded a "yes/no" response regarding readiness for hospital discharge, based on consideration of 6 items assessing behavioral and functional status. The RDQ was rated at Days 3, 5, 7, 14 and 21. Patients' satisfaction with their current treatment was evaluated using the Medication Satisfaction

Figure 6: Example page from annotated report