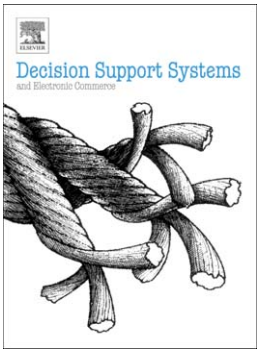
Accepted Manuscript



ADDIS: A decision support system for evidence-based medicine

Gert van Valkenhoef, Tommi Tervonen, Tijs Zwinkels, Bert de Brock, Hans Hillege

PII: S0167-9236(12)00260-6
DOI: doi: 10.1016/j.dss.2012.10.005
Reference: DECSUP 12168

To appear in: *Decision Support Systems*

Please cite this article as: Gert van Valkenhoef, Tommi Tervonen, Tijs Zwinkels, Bert de Brock, Hans Hillege, ADDIS: A decision support system for evidence-based medicine, *Decision Support Systems* (2012), doi: 10.1016/j.dss.2012.10.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

ADDIS: a decision support system for evidence-based medicine

Gert van Valkenhoef[a,b], Tommi Tervonen[*,c], Tijs Zwinkels[b], Bert de Brock[b], Hans Hillege[a]

[a] *Department of Epidemiology, University Medical Center Groningen, The Netherlands*
[b] *Faculty of Economics and Business, University of Groningen, The Netherlands*
[c] *Econometric Institute, Erasmus University Rotterdam, The Netherlands*

Abstract

Clinical trials are the main source of information for the efficacy and safety evaluation of medical treatments. Although they are of pivotal importance in evidence-based medicine, there is a lack of usable information systems providing data-analysis and decision support capabilities for aggregate clinical trial results. This is partly caused by unavailability (i) of trial data in a structured format suitable for re-analysis, and (ii) of a complete data model for aggregate level results. In this paper, we develop a unifying data model that enables the development of evidence-based decision support in the absence of a complete data model. We describe the supported decision processes and show how these are implemented in the open source ADDIS software. ADDIS enables semi-automated construction of meta-analyses, network meta-analyses and benefit-risk decision models, and provides visualization of all results.

Key words: Evidence-based medicine, evidence synthesis, data model, clinical trial, decision analysis

1. Introduction

Two kinds of decision support systems for evidence-based medicine can be distinguished: rule-based systems for supporting operational decisions of practising physicians and strategic decision support systems. The rule-based systems represent clinical knowledge and include inference rules for aiding professional decision making in clinical practice. They have been in existance since the 1970s [1]. The most common of these are Computerized Physician Order Entry (CPOE) systems which contain evidence-based rules that enable issuing warnings when an inappropriate combination of medicines is prescribed. To the best of our knowledge, there are no established systems that inform strategic (rather than operational) decisions such as identifying the best treatment practices based on the consideration of benefit-risk trade-offs.

Strategic health care decision making, with or without a supporting system, depends heavily on the availability of unbiased evidence from controlled clinical trials [2]. One of the core activities and sources of information in evidence-based medicine is the systematic review [3], a literature review that attempts to identify and synthesize all empirical

*Corresponding author: Econometric Institute, Erasmus University Rotterdam, PO Box 1738, 3000DR Rotterdam, The Netherlands. Tel.: +31 10 408 1260

Email addresses: g.h.m.van.valkenhoef@rug.nl (Gert van Valkenhoef), tervonen@ese.eur.nl (Tommi Tervonen), t.j.w.zwinkels@ai.rug.nl (Tijs Zwinkels), e.o.de.brock@rug.nl (Bert de Brock), j.l.hillege@epi.umcg.nl (Hans Hillege)

ACCEPTED MANUSCRIPT

evidence that fits pre-specified eligibility criteria in order to answer a specific research question [4]. Currently the process of systematic review is extremely labor intensive and error prone due to the lack of a comprehensive source of clinical trials, the inaccuracy of literature searches, interpretation issues, tedious manual data extraction and, importantly, the duplication of effort that is necessary for every review [5]. The emergence of clinical trial registries [6] and the move towards a more open clinical research community [7, 8], as well as the initiatives of the Cochrane foundation [9] to share and update meta-analysis data sets offer opportunities for more efficient approaches to evidence synthesis. Still, to date there is no single complete collection of performed clinical trials and outcome data, and importantly none of the available sources store results in a format that is suited for re-analysis [6, 10].

Thus, although suitable methods for evidence-based strategy decision support exist [11–14], evidence-based decision making is difficult to implement because of the substantial effort required to systematically review the literature for relevant studies and to manually extract the data from these studies, which has to be done on a case by case basis. Even when a relevant published systematic review exists, evidence-based decision making including multiple (possibly conflicting) objectives is difficult and in practice often done ad hoc due to a lack of supporting information technology. In addition, sometimes it will be necessary to incorporate additional studies to the body of evidence present in the systematic review, e.g. in the regulatory context where the manufacturer sponsors studies to prove the efficacy and safety of a newly developed drug. Moreover, the analyses reported in the published systematic review may not be valid for the decision at hand, so re-analysis of the included clinical trials may be needed. Text-based reports of systematic reviews do not support such use cases. There do exist methods for automated extraction of trial design and results from the literature, but although the field is rapidly evolving [see e.g. 15], their accuracy is not yet sufficient to be directly used in systems supporting strategic decisions.

In this paper, we present ADDIS (Aggregate Data Drug Information System, `http://drugis.org/addis`), an open source evidence-based drug oriented strategy decision support system. It is an integrated software application that provides decision support for strategic decisions such as guideline formulation, marketing authorization, and reimbursement. ADDIS stores aggregate clinical trial results with a unifying data model, and implements semi-automated evidence synthesis and benefit-risk modeling. These use cases were derived from direct discussion with experts from pharmaceutical industry, regulatory authorities, and academia, and from their feedback to early prototypes of the system. Before the models can be applied, trial results must be available in the system; for this, we present an assisted procedure for importing study designs from an existing database. The evidence synthesis and decision models of ADDIS allow decision makers to visualize and understand the available evidence and the trade-offs between different treatment options, thus addressing information overload and reducing the complexity of strategy decisions informed by clinical evidence. We stress that ADDIS does not aim at operational decision support, but aids in strategic decision making and provides a platform for computational methods in clinical trial informatics. In addition, the generation of the models cannot be completely automated: some steps require decisions from a domain expert, but can be supported by ADDIS as will be shown in this paper. To the best of our knowledge, ADDIS is the first system to allow on demand generation and use of the evidence synthesis and decision support models in a suitable way for strategic

ACCEPTED MANUSCRIPT

decision making.

We start by discussing existing systems and standards for clinical trial design and results in Section 2. The unifying data model is presented in Section 3. After that, in Section 4, we present ADDIS and the assisted procedures of study import and generation of evidence synthesis and benefit-risk models. In Section 5 we summarize our principal findings and propose directions for future research.

2. Background

Several systems and standards dealing with clinical trial information exist. We provide an overview of these systems and standards in Sections 2.1 and 2.2, respectively. Subsequently, in Section 2.3, we briefly describe the current state of methods for extraction of information from predominantly text-based sources of clinical trial designs and results. Finally, Sections 2.4 and 2.5 give an overview of the most relevant evidence synthesis and decision modeling approaches for strategic decision making.

2.1. Clinical trial information systems

In this section we briefly summarize the information systems that deal with clinical trials information, first those in operational management of trials and the regulatory environment, then the dissemination to the scientific community through publication in journals and registration, and finally how the results are summarized in systematic reviews.

2.1.1. Operational management and regulatory submission

Operational management refers to the administrative and data-gathering activities for a single trial. The operational management of clinical trials can be automated by using a Clinical Trial Management System (CTMS). Until circa 2000, the management and data collection of the vast majority of clinical trials were paper-based activities [16], but the use of a CTMS has quickly become the norm [17, 18]. The automation of operational management is now a mature field, and increasingly standardized (see also Section 2.2). However, CTMSs are data-centric single study systems that are focused on enabling the efficient operation of the trial and, often, submission of data to the US Food and Drug Administration (FDA). As of yet these systems do not enable cross-study analyses, data integration and data sharing.

After drug development, the pharmaceutical company compiles the evidence collected from clinical trials (and other research) into an electronic dossier that is submitted to the regulators who decide upon its market authorization. The dossier, especially the clinical trial results, forms the basis on which regulators assess the benefit-risk profile of a new drug. Submissions to the European Medicines Agency (EMA) and most other regulatory agencies worldwide are mainly text-based, containing aggregate-level results of clinical trials based on the applicant's statistical analyses. The FDA, on the other hand, requires an electronic submission of individual patient data to be able to perform independent analyses [19], and is currently building JANUS, a standards-based clinical data repository specifically designed for the integration of data [20].

ACCEPTED MANUSCRIPT

2.1.2. Results dissemination

Pharmaceutical companies and clinical research organizations may choose to publish the results of clinical trials in peer-reviewed scientific articles that do not include the underlying data set. Abstracts of publications are indexed in databases such as PubMed (`http://pubmed.com/`), which includes over 20 million citations from over 5,000 journals, of which more than 600,000 were published in 2009 [PubMed, 2011-05-02]. Although large in size, PubMed contains only a selected subset of the biomedical literature [21]. Abstract databases include meta data that might be incomplete due to being provided by external parties; for example, to achieve high sensitivity in searching for clinical trials in PubMed, restricting the search to the 'clinical trial' publication type is too restrictive [22], and a broader query is recommended [4]. The Cochrane CENTRAL database of clinical trials is dedicated to indexing reports of clinical trials only, and contains references to 645,086 publications of clinical trials, of which 286,418 have been published since 2000 [Cochrane Library, 2011-05-02].

Until recently journal publications were the only non-confidential source of trial designs and results. This led to insufficient or inaccurate trial reporting and publication bias [23] as e.g. over half of the clinical trials supporting successful new drug submissions made to the FDA had still not been published 5 years after the medicines' market approval [24]. Publication bias is a serious problem that can lead to incorrect conclusions in a systematic review. As early as in 1986 the registration of trials in advance was proposed as a solution to publication bias [25]. In 1997 the US became the first country to make trial registration a legal requirement, leading to the development of the ClinicalTrials.gov registry [26]. In 2004, both the World Health Organization (WHO) and the International Committee of Medical Journal Editors (ICMJE) released statements in support of the prospective registration of clinical trials. This policy has been widely adopted [27] and now assures that the existence of in any case most (recent) trials is known [6]. Registries primarily focus on providing a record of trials for enabling patient recruitment and investigator accountability. Various organizations, including the WHO, have called for a full disclosure of the trial protocol (including amendments) and results [7, 8, 28–32], but only the US have adopted legislation that requires registering results in ClinicalTrials.gov [10, 33]. Study protocols can be retrieved from ClinicalTrials.gov in a (semi-structured) XML format [34], while the retrieval of results is only possible in a text-based format. Other registries provide protocol information as semi-structured text, and do not include results.

In order to unify trial registration worldwide the WHO Registry Network was established in 2007. Twelve national and international registries are now part of the network. The European Union clinical trials registry, EudraCT, was opened to the public only recently, on 22 March 2011 [35], and is not part of the WHOs Registry Network. Table 1 gives an overview of the WHO primary registries, ClinicalTrials.gov, and EudraCT. ClinicalTrials.gov is by far the largest registry, containing more than 8 times the number of trials recorded in the second largest registry (EudraCT).

2.1.3. Systematic review

Evidence-Based Medicine (EBM) tries to use the best available evidence in assessing the benefits and risks of a treatment [2]. The most frequently implemented methods to assess the available evidence are the systematic review

ACCEPTED MANUSCRIPT

Register	Studies	Results
ClinicalTrials.gov (United States)	106,649	yes (3,441)
EudraCT, the *European Union* Clinical Trials Register	12,990	no
ISRCTN register (international)	9,645	no
Japan Primary Registries Network	6,193	no
Australian New Zealand Clinical Trials Registry	5,221	no
The Netherlands National Trial Register	2,728	no
Clinical Trials Registry - *India*	1,704	no
Chinese Clinical Trial Register	1,319	no
Iranian Registry of Clinical Trials	1,291	no
German Clinical Trials Register	482	no
South Korea Clinical Research Information Service	108	no
Cuban Public Registry of Clinical Trials	105	no
Sri Lanka Clinical Trials Registry	60	no
Pan African Clinical Trial Registry	48	no

Table 1: ClinicalTrials.gov, EudraCT, and the 12 WHO primary registries. The 'studies' column indicates the number of registered trials (per 2 May 2011) and the 'results' column whether the registry also enables results publication.

and meta-analysis of published research results [3]. Systematic reviews are usually presented in a textual format without the underlying dataset. Given the effort required to perform a systematic review, fragmented reports regarding an indication are common [36]. The rapidly growing number of systematic reviews published each year [37] has led to the 'overview of reviews' or 'umbrella review' to summarize the results of the existing reviews for an indication [38]. Umbrella reviews generally merely repeat the pooled summaries of treatment effects from the original reviews, but it has been argued that they may lead to misleading and inconsistent conclusions [36].

The flagship of systematic reviewing is the Cochrane Library, kept up-to-date by the non-commercial Cochrane Collaboration (`http://www.cochrane.org/`). It is composed of three main components: the CENTRAL literature database of trial publications [39], the Cochrane Database of Systematic Reviews [40], and software for conducting and reporting on meta-analyses (`http://ims.cochrane.org/revman`). The Cochrane Library provides reviews of effects of healthcare interventions generated and updated by medical researchers, which are on average regarded to be of better quality than the corresponding studies published in traditional journals [41]. Compared with the traditional journal publications that usually provide data in tables or figures, the Cochrane Reviews incorporate descriptions and results of the original studies, while the software enables making odds-ratio diagrams that can also include the newest studies. However, the available datasets are not complete and they are structured according to the reviews rather than the included studies. Moreover, due to the inaccessibility of clinical trials information, systematic reviews are static

ACCEPTED MANUSCRIPT

entities that only reflect the state of knowledge at the time of the literature search.

2.2. Standards and data models

The information systems discussed above, especially those in operational management, are enabled by standards and data models that have been developed over the last two decades. Two main standardization bodies in the field are the Clinical Data Interchange Standards Consortium (CDISC) and Health Level 7 (HL7). The CDISC develops vendor-neutral and freely available standards that enable information system interoperability in the operational management and regulatory submission of clinical trials. HL7 develops standards that apply broadly to clinical and administrative data in health care, and thus do not focus on any specific clinical domain. The foundation of HL7 standards development work is the Reference Information Model (RIM), a high level object model of the health care domain. Several standards are derived from the RIM, such as V3 Messages for the meaningful interchange of data between health care systems, GELLO for rule-based decision support, and the Clinical Document Architecture for semantically structured documents. HL7 also maintains the Arden Syntax that enables rule-based expert systems that support operational decision making in health care.

The Biomedical Research Integrated Domain Group (BRIDG) project aims at bringing together the common elements of their various standards to a shared view of semantics of the domain of protocol-driven research and its associated regulatory artifacts [42]. The model is intended to be implementation independent in the sense that it models the problem domain, and not any specific solution. For example, unlike some other CDISC standards it does not specify the format in which to submit data to the FDA. The BRIDG model is subdivided into several sub-domain views: the protocol representation, study conduct, adverse event and regulatory perspectives. While the operational aspects of clinical trials are well covered by these perspectives, a data analysis perspective is currently missing as there is no adequate standard for statistical analysis.

The ClinicalTrials.gov registry has developed their own model, the Data Element Definitions (DED) [43, 44]. They allow the reporting of aggregated outcome data and statistical analyses to some extent, but the semantic depth of the information is limited as most fields are free text. For example, since eligibility criteria are free text fields, searching for a trial relevant to a specific patient condition is inaccurate [45].

The Human Studies Database (HSDB) project aims to share fully machine understandable representations of study design information between institutions [46]. To enable this, they develop the Ontology of Clinical Research (OCRe), which defines the concepts that should be queried across the individual institutions' databases. The creators of OCRe have argued that while the BRIDG model accurately captures the operational semantics of clinical trials, its modeling of many aspects relevant to cross-study analyses is weak [46]. The main contributions of OCRe at this time are a study design topology [47], the ERGO formal machine readable representation of eligibility criteria [45], and a model of study outcomes that separates the phenomena of interest from the variables that code them [46]. It also contains a study design representation derived from BRIDG [46]. While OCRe is a promising effort, it is still far from comprehensively representing study design and lacks results completely.

ACCEPTED MANUSCRIPT

Model	Study design	Aggregate results	Semantic depth	Completeness
BRIDG	++	-	+/-	+
DED	+/-	+/-	-	+
OCRe	+	-	+	-
OBX	+	+/-	+/-	+

Table 2: Approximate scoring of data models on several dimensions relevant to automated processing of aggregate clinical trials results. 'Study design' is the extent to which complex study designs can be accurately represented, 'aggregate results' refers to the inclusion of aggregate results and description of the means by which they were derived, 'semantic depth' refers to the level of semantic structure achieved (e.g. contrast the text-based eligibility criteria in the DED to the ERGO model used in OCRe), while 'completeness' refers to the extent to which the model in its current state achieves its stated goals. Note that these dimensions are difficult to assess and the assigned ratings are subjective.

The Ontology Based Extensible Conceptual Model (OBX) is another ontology for representing clinical trials [48, 49]. It is specifically aimed at making available the results of immunology studies for data re-use and re-analysis. The OBX also incorporates study design representation ideas from BRIDG and the ClinicalTrials.gov DED [48]. While it appears successful in developing a broadly applicable data model for biomedical studies, and also includes results, it would appear that the objections raised by HSDB researchers about the depth of modeling in BRIDG also apply to OBX, and the results are represented in a way similar to the ClinicalTrials.gov DED.

We rate the four major models in Table 2 on how well they represent study design and aggregate results, as well as their semantic depth and completeness. Table 3 gives a summary of the main goal as well as the strengths and weaknesses of each model. One common property of all models is that they rely on an external terminology for their clinical content. Controlled terminologies (synonymously: controlled vocabularies, coding systems) of clinical terms are an important first step in the application of information technology to medicine [50]. Controlled terminologies predate information technology, e.g. the International Classification of Diseases (ICD) was already introduced in 1893. The ICD formally codes diseases and enables, for example, the assessment of disease incidence from medical records. Other terminologies fill other niches, for example the Medical Subject Headings (MeSH) [51] is used to index the medical literature (e.g. PubMed meta data is coded in MeSH), and the Medical Dictionary for Regulatory Activities (MedDRA) is used for coding safety data (e.g. adverse events). Many of these specialized terminologies are organized into a strict hierarchy, which means that some specific terms may fit in multiple places [50]. The Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT) terminology is an important attempt to create a clinical terminology with comprehensive coverage [52]. It currently contains around 311,000 concepts and 800,000 terms [53]. It also goes beyond a simple hierarchical structure and provides the logical relationships that hold between terms; over 1.3 million such relationships are currently modeled [52, 53]. The Unified Medical Language System (UMLS) [54, 55] 'Metathesaurus' brings together over 60 biomedical terminologies and their relationships. The ICD, SNOMED CT and MeSH are among the integrated terminologies.

ACCEPTED MANUSCRIPT

Model	Purpose	Strengths	Weaknesses
BRIDG	Operational management, regulatory submission	Standardization process, practical applications	No aggregate data, limited depth of modeling (some aspects)
DED	Limit publication bias, enable disclosure of results	Completeness, working system, US trials required to register (data available)	Limited semantic structure
OCRe	'Computable' representation of human studies	Broad scope, semantic depth	Not finished, not implemented, results not represented
OBX	Make available data from immunology studies	Balance of the depth of modeling and the practical feasibility, working system	Limited depth of modeling (some aspects)

Table 3: The purpose of each of the data models as well as their strengths and weaknesses from the perspective of enabling automated evidence synthesis and decision support.

2.3. Data extraction

The free-text nature of clinical trial publications is an important obstacle to the application of data mining and other automated knowledge discovery and decision aid uses [5]. There are many existing approaches to extract some of this data from abstracts or full texts of journal articles or health records, as reviewed in [56, 57]. However, the potential benefits are currently not fully realized due to lack of directly applicable tools [56] and text mining approaches for supporting research [57].

Text mining of articles describing clinical trials could support researchers in performing a systematic review. Information extraction, on the one hand, attempts to create structured datasets from unstructured text by identifying entities and relationships between entities in the text. Most current approaches focus on the abstract rather than the full text as it provides a more controlled environment, and they tend to focus on only a few information elements [15]. The ExaCT system [15] assists systematic reviewers in extracting 21 key trial characteristics from full text articles. The system is accurate enough to save a considerable amount of time in extracting these elements, but systematic reviewers do have to verify the extracted information manually. Text analytics, on the other hand, identifies patterns in large collections of texts in order to classify documents and unlock relationships between documents. Text analytics can help systematic reviewers in structuring large sets of search results from abstract databases (e.g. PubMed) and increase the efficiency of finding the relevant clinical trials. However, to be able to reliably perform evidence synthesis and decision modeling based on the extracted clinical trials data, a higher level of accuracy and generality is needed than is currently offered by text mining methods. Thus, although automated methods can lower the workload, manual

ACCEPTED MANUSCRIPT

data extraction remains necessary.

2.4. Evidence synthesis

The most commonly applied evidence synthesis method is pair-wise meta-analysis, in which a number of studies comparing the same pair of treatments A and B are synthesized to assess their relative performance δ^{AB} on a specific outcome [58]. For example, do more depressed patients respond to treatment with paroxetine than with fluoxetine (both antidepressants)? Or, do more patients treated with paroxetine experience nausea during the studies than those treated with fluoxetine? The observed treatment differences $\hat{\delta}_i^{AB}$ in the individual studies i are used to estimate the overall difference δ^{AB}. Network meta-analysis, a recent extension of pair-wise meta-analysis, synthesizes evidence on the relative effects of a whole network of treatments simultaneously [59–61]. It incorporates both direct and indirect evidence on the relative effects, and allows a statistical analysis of evidence consistency [61–63]. Except for the possible inconsistency between direct and indirect evidence, the assumptions underlying network meta-analysis are the same as those underlying pair-wise meta-analysis [64]. The method has gained acceptance, and applications are being published in top medical journals [e.g. 65, 66]. However, application of the method has so far remained the work of a select few experts, as model specification is difficult and no automated tools are available. Many other evidence synthesis methods exist [67], but pair-wise and network meta-analysis are by far the most important ones for decision support.

2.5. Decision models

Although evidence synthesis is an important tool for evidence-based medicine as it helps to summarize the available evidence, it does not help the decision maker to take into account the trade-offs of the risks of a treatment and its related benefits. There is an increasing interest in evidence-based multi-criteria decision models [68, 69] taking into account efficacy and safety of alternative treatments. The target domains of model-based decisions include marketing authorization for new drugs, development of guidelines concerning recommended treatments, and prescription decisions such as which anti-depressant to subscribe, for example, in a setting where besides efficacy specific safety issues are also of interest, e.g. dizziness could be life-threatening given the specific patient's occupation. Many such decisions have to take into account trade-offs between different decision criteria (e.g. efficacy and safety), and can be aided through multi-criteria decision models [70] or application-specific ways of mapping benefits and risks to a single scale [71]. Multi-criteria decision models can structure the decision problem and make trade-offs between the alternative medical treatments explicit. In general, Multi-Criteria Decision Analysis (MCDA) methods compare m alternatives on n criteria. The performance of each of the alternatives is measured in terms of the criteria, and explicit trade-offs (preferences) between the criteria may be specified by the decision maker. The decision is aided by finding the optimal alternative (choice problem), by ranking the alternatives from best to worst, or by classifying the alternatives into discrete classes, such as good, acceptable and bad alternatives [72]. An inverse approach, in which typical preferences that favor each of the alternatives are derived using the decision model, is also possible [73].

ACCEPTED MANUSCRIPT

There exists benefit-risk models based on point estimates of the criteria measurements [13, 74]. However, taking into account decision uncertainty is necessary in the medical context as the data might not distinguish the alternatives with sufficient certainty to make an informed decision. In that case, a decision has to be postponed until more or higher quality information becomes available. Therefore, we focus on stochastic methods, where the performances are measured using probability distributions rather than with point estimates. Stochastic methods based on single studies [11, 75] model the 'absolute' treatment effects and use those as performance measures (e.g. the binomial success probability of a treatment response can be modeled using a Beta distribution for each of the treatments). Using absolute measures has the advantage that the observed differences in performance have an immediate clinical implication, and thus eliciting preferences from the decision maker is relatively easy. For example, one could ask 'Would you consider improving the probability of treatment response from 0.73 to 0.80 to be more important than reducing the probability of the side effect dizziness from 0.12 to 0.09?'. However, the generalizability of a model using absolute measures is questionable, as the absolute treatment effect and the incidence of side effects depends heavily on the design and specific population of the study. Models based on evidence synthesis [14] are preferable from this perspective, as measurements would be based on *relative* effects estimated using all available studies. Thus, such a method is more robust and generalizable, but the relative scales make the interpretation of the clinical implications more difficult [76]. A hybrid approach, in which the (relative) measurements are derived using evidence synthesis, but framed in clinically meaningful (absolute) terms using (assumed or estimated) baseline risk for the population of interest may be the best one [14].

So far, only benefit-risk models based on Stochastic Multi-criteria Acceptability Analysis (SMAA) [73, 77, 78] allow taking into account the full uncertainty surrounding the measurements from clinical trials as well as imprecise preferences, while enabling the comparison of $m \geq 2$ treatments on $n \geq 2$ outcomes through Monte Carlo simulation. A two-dimensional visual approach (also based on Monte Carlo simulation) may be preferable if $m = 2$ and $n = 2$ [75]. This model is based on standard cost-effectiveness analysis techniques, and we shall refer to it as the "Lynd & O'Brien" model. Both methods enable the inverse approach, where the preferences supporting specific decisions are derived using the model.

3. The unifying data model

We developed a unifying data model to enable evidence-based decision support methods based on either individual studies or evidence synthesis. As discussed before, the most important methods are pair-wise meta-analysis, network meta-analysis, and stochastic multi-criteria benefit-risk assessment. The data model is aimed at supporting these use cases. As was shown in Section 2.2, several worthwhile data modeling efforts are underway. Unfortunately none of them have the needed level of modeling to be directly applicable to our use cases. It is clear that while very precise representations (such as are being created for OCRe) will be needed for the reuse of clinical trial data for general purposes, they will not translate directly to the application of evidence synthesis.

ACCEPTED MANUSCRIPT

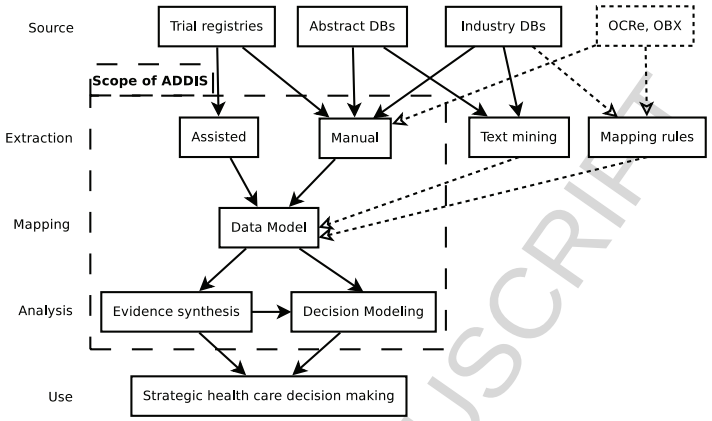


Figure 1: The current and future flow of information into the ADDIS system, and the role of the unifying data model in supporting evidence synthesis and decision support. The dashed rectangle indicates the scope of the functionality currently implemented by ADDIS. The solid arrows show the current situation, while dotted arrows indicate how future developments will benefit ADDIS.

In addition, the data model is intended to be a well defined data extraction target to enable health strategy decision support. Data extraction is currently based on manual and assisted import from clinical trial registries and journal publications, and in the future on automated rule-based import from structured databases or semi-automated data extraction from the medical literature. This vision is shown in Figure 1. The data model is not intended to fully model clinical trials, as we believe BRIDG and OCRe are better positioned to eventually fill this gap. Rather, there is a need for a unifying data model that captures the invariants of the domain from the perspective of evidence synthesis. Such a data model provides clear requirements for more fine-grained models, a target for text mining and (sub-)domain-specific rules for data conversion, and a basis on which to build decision support systems. Thus, our data model represents the structure of trials only to a limited extent and appropriate (domain-specific) mapping is required to enable its use. Mapping rules from more fine-grained data models such as OCRe can be developed once these models have matured. The unifying data model is described below and illustrated in Figure 2. In the text, we will refer to entities in the domain model using capitalized words (e.g., Study and OutcomeMeasure).

Clinical trials are represented by the class Study. The data model may also apply to other studies with human populations, such as observational studies, but it was primarily designed to represent clinical trials. Each Study is identified by a name (e.g., "Coleman et al. 2001" or "NCT00296517"). A Study considers a single (therapeutic) Indication. Each Indication is identified by a definition (e.g., "Depression" or "Type 2 Diabetes"). A Study consists of (two or more) Arms. An Arm within the context of a clinical trial can be seen as a group of patients within a Study who all receive the same medical Treatment. Within a Study there can exist different Arms for the same medical Treatment (e.g., receiving different dosages). Each Treatment is identified by a definition (e.g., "Placebo", a simulated medical intervention, "Fluoxetine", an anti-depressant, or "Rosiglitazone", an anti-diabetic).

An OutcomeMeasure is identified by a definition, referring to an endpoint (e.g., "Responders on the HAM-D

ACCEPTED MANUSCRIPT

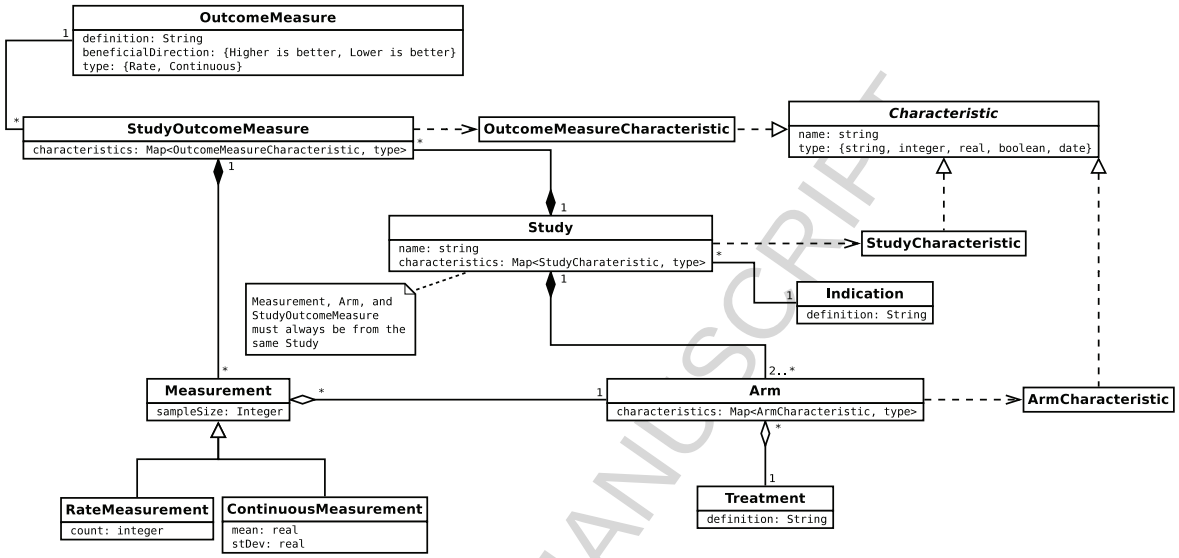


Figure 2: The unifying data model for common types of aggregate analysis of clinical studies in UML2 class notation.

rating scale" or "Change from baseline triglycerid levels (mg/dL)") to be measured in studies, or an adverse event (e.g., "Headache", "Nausea" or "Chest pain") that can occur in studies. An OutcomeMeasure has a beneficial direction (higher is better or lower is better). There are two Types of OutcomeMeasures in terms of how they are measured: rate or continuous (see below). A Study can have (zero or more) OutcomeMeasures and an OutcomeMeasure can apply to (zero or more) Studies. Such a combination is called a StudyOutcomeMeasure (identified by the OutcomeMeasure and Study). Note that the BRIDG model discussed before also uses the term StudyOutcomeMeasure in this context.

A Measurement refers to a combination of a StudyOutcomeMeasure and an Arm within the same Study. Each such combination can have at most one Measurement. A Measurement has a sample size (e.g. 98 patients). The sample size is associated with the Measurement and not with the Arm, as the relevant sample size depends on the way the outcome measure is analyzed, and may change over time due to patients dropping out of the study. Each Measurement is either a RateMeasurement or a ContinuousMeasurement, depending on the type of OutcomeMeasure the measurement refers to. A RateMeasurement describes the number of individuals in the Arm for whom the OutcomeMeasure occurred. A ContinuousMeasurement describes the result by a mean and a standard deviation (two real numbers).

The entities described above form the core of the unifying data model. Generation of evidence synthesis and decision models is based on Studies, Arms, Treatments, OutcomeMeasures, and Measurements. In addition, the data model includes Characteristics for more descriptive information. The Characteristics are identified by a name (e.g. for StudyCharacteristics "Study size", "Group allocation", "Treatment blinding" or "Patient eligibility criteria", for ArmCharacteristics "Arm size", "Dosing" or "Gender distribution", for OutcomeCharacteristics "Is primary outcome" or "Assessment time") and include the type of the characteristic. The type is used for validating the values input for the actual characteristic values, and is useful in generating Graphical User Interface (GUI) components for input of

ACCEPTED MANUSCRIPT

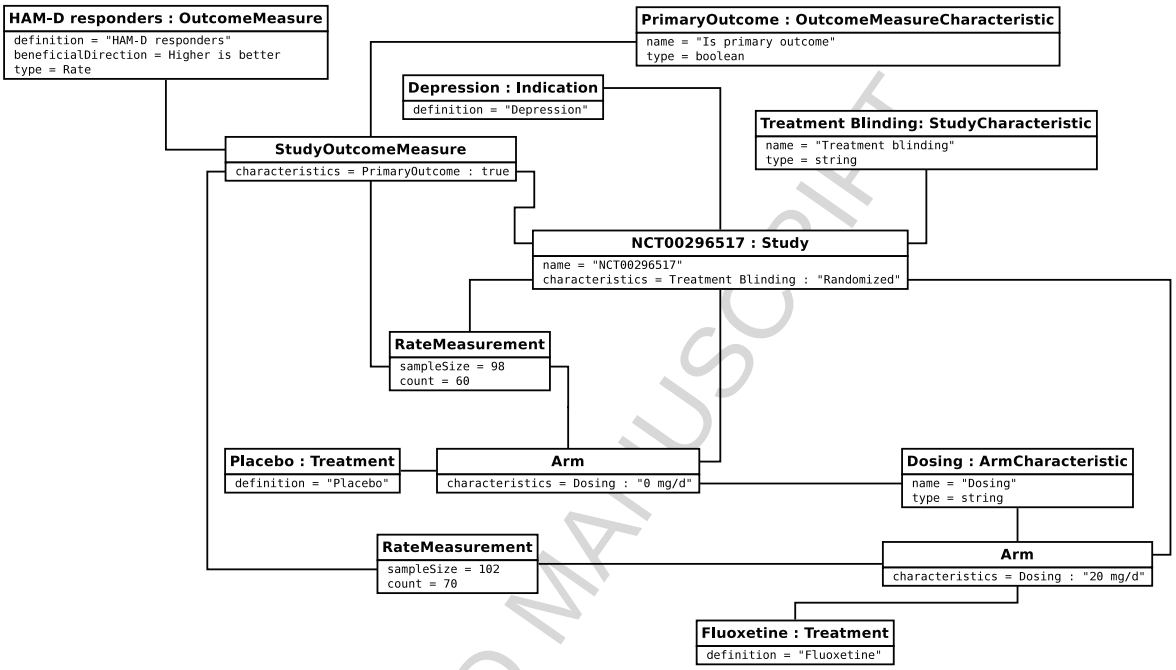


Figure 3: An example instantiation of the unifying data model as UML2 object diagram. The example instantiation depicts one Study including two Arms with two different Treatments. For both of the arms, a measurement on one OutcomeMeasure is shown. For the Arm, StudyOutcomeMeasure, and Study, there is each one Characteristic presented together with the associated value.

characteristic values. The object diagram in Figure 3 includes an example instantiation of the data model.

If the characteristics are left out, the data model contains the minimal information for generation of the evidence synthesis and decision models described in the previous section. The minimal representation makes it easier to import data to a system implementing it, and increases applicability of the model from being specific to a certain subfield (e.g. cancer treatments) to being general for all. However, the minimality causes the data model to be specific for the chosen types of analysis models. If, for example, meta-regression techniques should be applied, the data model would need to be extended accordingly. We allow descriptive extensions by including the characteristics. The characteristics also serve for storing the information that is unnecessary for analysis model generation, but necessary for expert judgment on which studies should be included in the analysis (e.g. based on the type of dosing). They also serve for specialization of the data model in that, if the need arises, new ones can be added without breaking the functionality of analysis model generation.

ACCEPTED MANUSCRIPT

4. ADDIS decision support system

The unifying data model together with a semi-automated analysis generation system are implemented in the open source decision support software ADDIS[1]. It provides an easy interface to enter, import and manage study design and outcome information from clinical trials, and is specifically aimed at supporting the user in creating (network) meta-analyses and (multi-criteria) benefit-risk models. The main components of the software are:

- Implementation of the unifying clinical trial data model,
- GUI for managing trials and analyses,
- Semi-automated import of studies from ClinicalTrials.gov,
- GUI 'wizards' for semi-automated generation of analyses,
- External packages for computing the analyses,
- GUI components for results visualization, and
- Links to external databases (PubMed, ATC database, drug compendium).

ADDIS integrates an external network meta-analysis library[2] [79] and JSMAA [80] for computation of SMAA benefit-risk models. The ADDIS data format is represented by an XML schema[3] that instantiates the unifying data model. Evolution of the format is supported by versioned XML schemas that are forward-compatible through XSL transformations (XSLT). ADDIS supports the coding of drugs with their Anatomical Therapeutic Chemical Classification System (ATC) code and uses them to link to drug compendia. The ATC codes can be filled in automatically by ADDIS through integration with an online database, when given the compound name. A coding system for outcome measures and adverse events will be integrated in the future. A number of study characteristics are supported by default in ADDIS to enhance the user experience. For studies, these include the study title, randomization, treatment blinding, the study objective, the in- and exclusion criteria, the start and end date of the study, PubMed IDs of relevant publications and several others.

4.1. Study import from ClinicalTrials.gov

The ClinicalTrials.gov registry is by far the most comprehensive clinical trials registry in the world, currently containing information on over 100,000 trials (see Table 1). ClinicalTrials.gov has a simple and easy to use interface to programmatically search for trials and retrieve their protocols in XML format (according to their own DED) [34]. Unfortunately, the results are currently not available as XML and it is unclear when this will be remedied.

[1] `http://drugis.org/addis`. Screen casts are based on version 1.6, `http://drugis.org/addis1.6`.
[2] `http://drugis.org/mtc`
[3] `http://drugis.org/files/addis-1.xsd`

ACCEPTED MANUSCRIPT

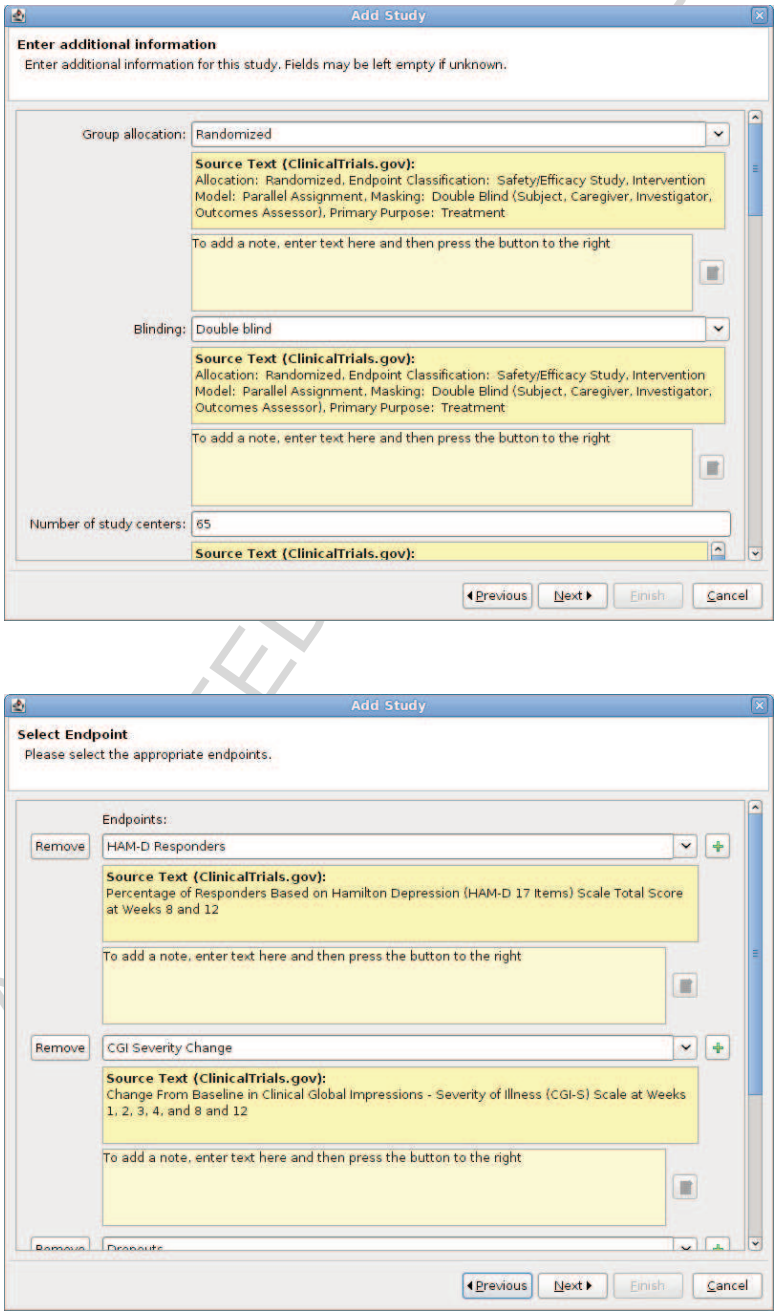


Figure 4: Example screens from the study input/import wizard. The top screen shows how study characteristics are input. Most of these are matched automatically from the source text. The bottom screen shows that endpoints must be mapped to entities in the database by the user.

ACCEPTED MANUSCRIPT

In ADDIS, we use this XML interface to semi-automatically import studies from ClinicalTrials.gov. The user inputs the NCT-ID of the trial that should be imported, and the software will retrieve the XML, from which it automatically fills in fields. For example, many of the study characteristics, such as randomization and treatment blinding are matched from DED fields using simple rules. However, those fields that form the core of our data model, such as the indication, treatments and outcome measures, have to be manually mapped to entities in the database. This is so because accurately mapping the free-text descriptions given in the ClinicalTrials.gov records would require (1) deep semantic modelling of the entities in our database, and (2) natural language processing of incredibly high accuracy. While both fields are rapidly evolving, neither of these problems have a fully satisfactory solution at the moment. This mapping step is critically important to the correctness of subsequent analyses and thus inaccurate automatic mapping could degrade the decision makers' trust in the system. Hence, for the time being, the mapping is deliberately left to the user. Figure 4 shows examples of the user interface for study import. The original source text is preserved as a note that is kept with the relevant field, and the user can also enter additional notes. Due to the lack of an XML interface for study results, those have to be entered manually, and can not be linked to the source text.

4.2. Evidence synthesis

ADDIS assists generation of pair-wise and network meta-analyses in a step-wise fashion; the process is presented in Figure 5. To start, the user needs to select an indication. Based on the selected indication, the system selects and presents all outcome measures included in the different available studies in the system considering the indication. After the user selects the desired outcome measure for analysis, the system selects the studies and their included treatments based on the selected (indication, outcome measure) tuple, and constructs the evidence graph. The graph is presented visually and has the vertices labelled with treatment definitions and the edges labelled with the number of studies including that comparison (see Figure 6). The user can pick the treatments to be compared. For a pair-wise analysis, exactly two treatments have to be selected, and for network meta-analysis two or more treatments can be selected. The software will not allow the user to continue unless the selected treatments form a connected graph. Following the treatment selection, the system presents the set of studies together with their characteristics, and non-desired studies can be easily removed by the user on a case by case basis. The chosen treatments and studies must form a connected evidence graph. Finally, if studies include a specific treatment in more than one arm (e.g. in various doses), the user must choose which arm to use in the analysis (see Figure 7).

Visualization of results is of crucial importance for applicability of methods used in evidence-based medicine. ADDIS provides visualization of the odds ratios, mean differences, risk ratios, and risk differences of standard meta-analyses in terms of forest plots (Figure 8). The network meta-analysis rank probabilities are presented as bar charts as shown in Figure 9.

4.3. Benefit-Risk models

The creation of benefit-risk models in ADDIS can be based on either an individual study or (previously created) meta-analyses. The user first selects an indication, and chooses whether to base the analysis on a single study or

ACCEPTED MANUSCRIPT

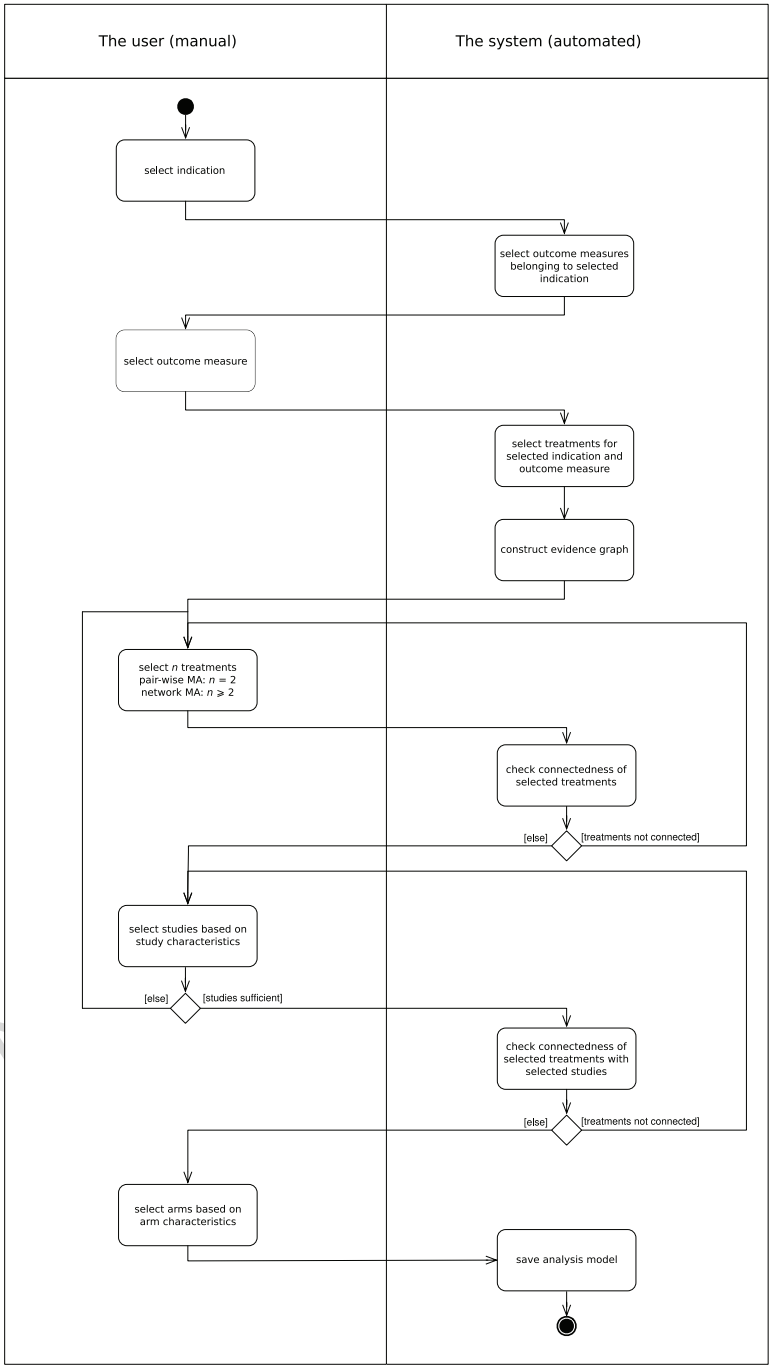


Figure 5: The process of meta-analysis creation as an activity diagram. The activities on the right-hand side are automated in the system, and the steps on the left require input and conscious decisions from the user. The process is identical for pair-wise and network meta-analysis, except that for pair-wise meta-analysis the number of treatments is restricted to exactly two.

ACCEPTED MANUSCRIPT

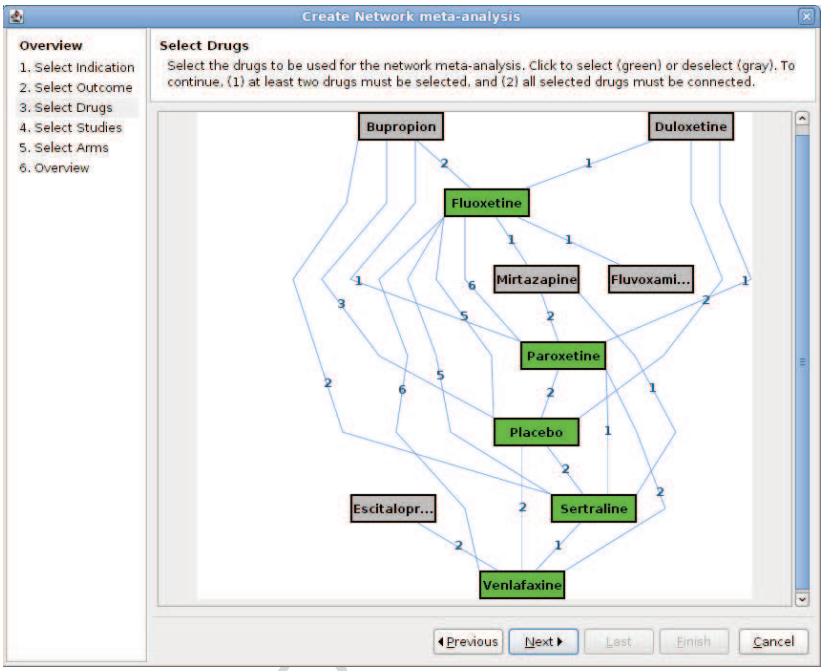


Figure 6: An evidence graph for a network meta-analysis. The treatments are the vertices, and the number of studies for each comparison label the edges (e.g., six studies compare fluoxetine and paroxetine). The green treatments are included in the analysis, the grey ones excluded.

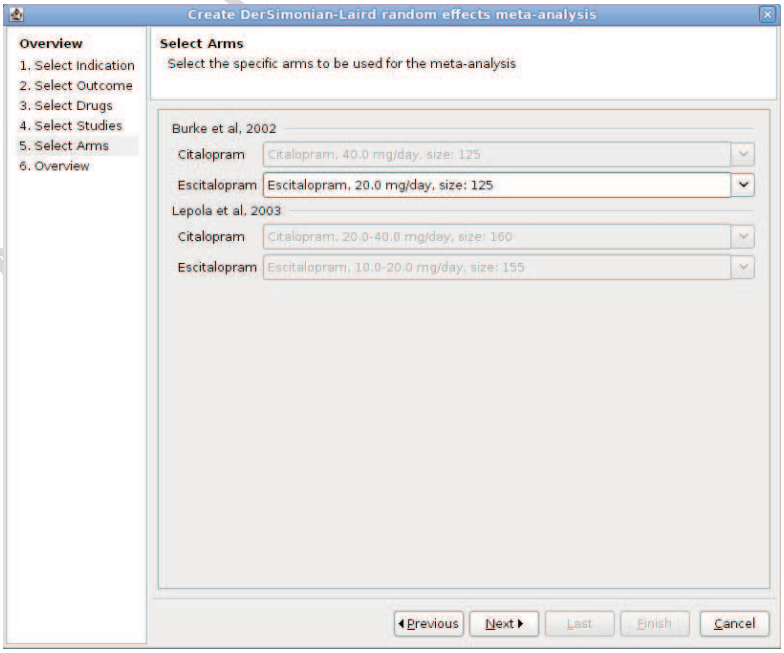


Figure 7: If several matching arms are available, the user must select an appropriate one based on the arm's characteristics. If the available arms are not appropriate, the user can go back to exclude the study.

ACCEPTED MANUSCRIPT

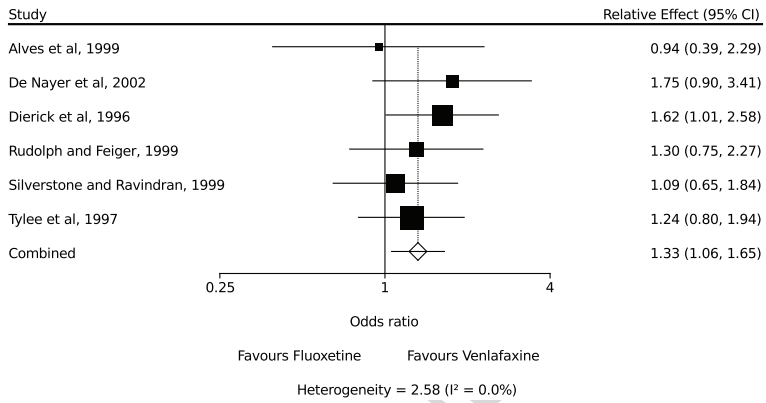


Figure 8: Visualization of standard meta-analysis results as a forest plot [81]. Here, odds-ratios (95% confidence intervals) are plotted on a logarithmic scale, with the pooled estimate shown last.

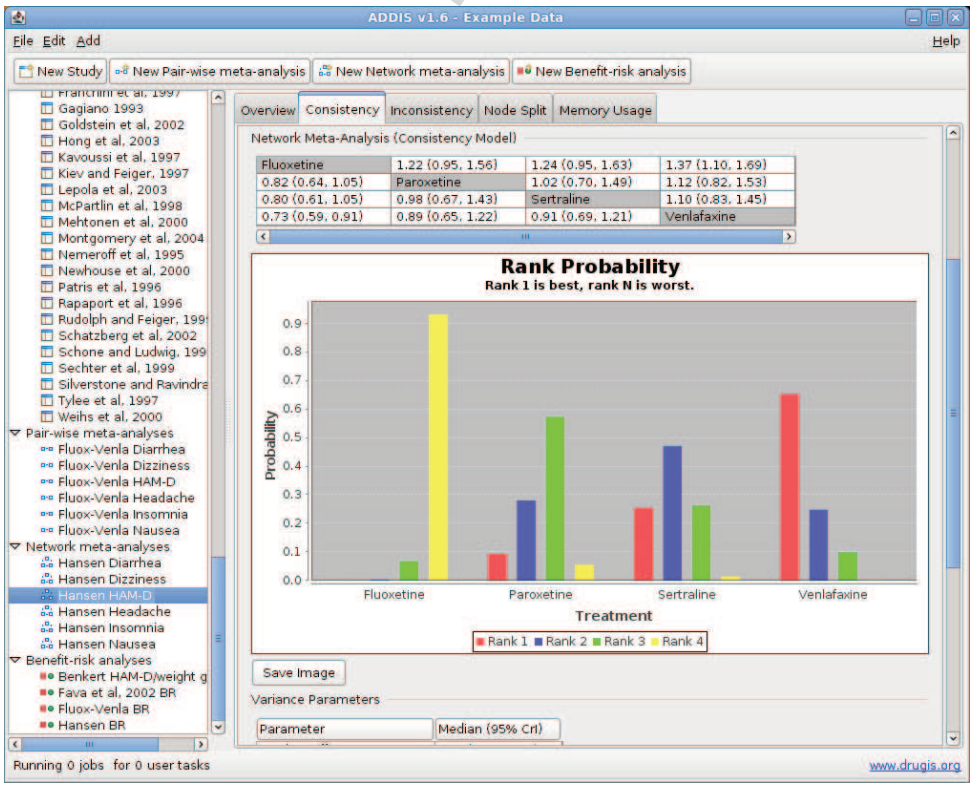


Figure 9: Network meta-analysis results. The table gives (posterior) odds-ratios (95% credibility interval) for all treatments relative to each other. The bar chart visualizes the (posterior) probability for each treatment to be best, second-best, etc. given the analysis model and the data.

ACCEPTED MANUSCRIPT

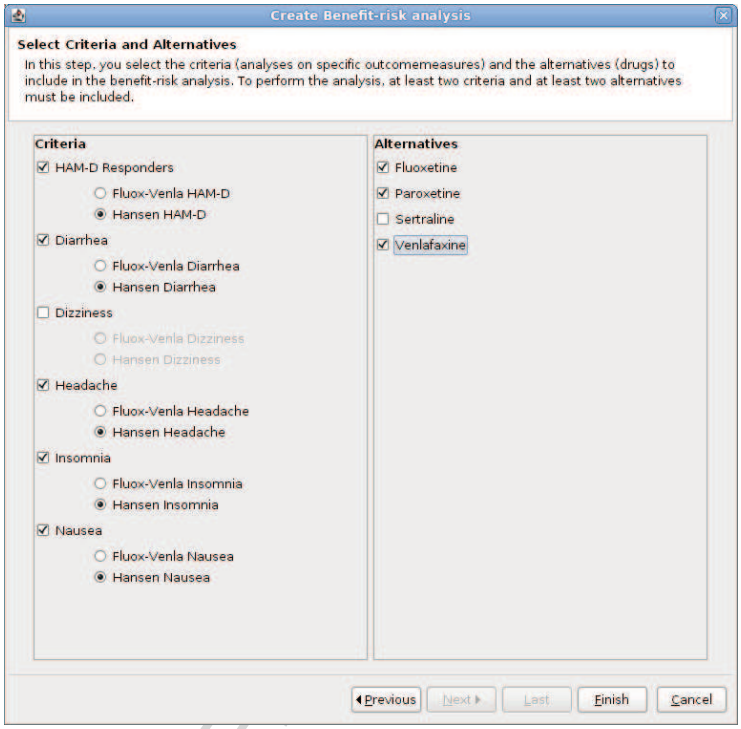


Figure 10: Criteria selection screen for construction of a benefit-risk model with synthesized evidence.

evidence synthesis. If the analysis is based on a single study, the system selects studies belonging to the selected indication and allows the user to choose one. Then, the user is presented with the available criteria (outcome measures in the selected study) and alternatives (arms in the selected study), and may select two or more of each to include in the benefit-risk model. If the analysis is based on evidence synthesis, the available criteria are the outcome measures for which a (network) meta-analysis exists within the selected indication. If multiple analyses exist for an outcome measure, one must be chosen. The available alternatives are the intersection of the sets of treatments included in the selected analyses. Two or more criteria and alternatives can be selected to include in the benefit-risk model. The final step in the creation of a benefit-risk model based on evidence synthesis is shown in Figure 10.

Benefit-risk decision models were already broadly discussed in Section 2.5. ADDIS supports decision makers using several different methods (see Table 4). These methods are organized along three axes: the number of alternatives, the number of criteria and the number of clinical trials in the evidence base. For a single-criterion decision between two alternatives based on a single study, standard statistical methods are sufficient. When there are several studies, pair-wise meta-analysis [58] can be used to pool the evidence, and for more than two alternatives network meta-analysis is needed [62]. When two criteria (e.g. one benefit and one risk) and two alternatives are to be considered, the "Lynd & O'Brien" model [75] based on either a single study or two meta-analyses (one for each criterion) can be used. For more than two alternatives or criteria Stochastic Multi-criteria Acceptability Analysis (SMAA) based

ACCEPTED MANUSCRIPT

criteria / **treatments**	1	2	≥ 2
2	PMA	L&O (S/PMA/NMA)	SMAA (PMA/NMA)
≥ 2	NMA	SMAA (S/NMA)	SMAA (S/NMA)

Table 4: Supported methods. Abbreviations: S = Single-study, PMA = Pair-wise meta-analysis, NMA = Network meta-analysis, L&O = Lynd & O'Brien benefit-risk, SMAA = SMAA-based benefit-risk.

models are available [11, 14]. The SMAA methods used are described in [73, 77] and their computational details in [82]. These can be based on either a single study, pair-wise meta-analyses (limited to 2 alternatives), or network meta-analyses (for ≥ 2 alternatives).

The results of a "Lynd & O'Brien" benefit-risk analysis are visualized both through plotting points from the probability distributions of incremental benefit and risk on the benefit-risk plane, and through the benefit-risk acceptability curve [75]. The SMAA models are visualized using the JSMAA visualization components and tables, showing the rank acceptabilities (Figure 11) to indicate how likely the alternatives are to obtain a certain rank (from best to worst) and the central weights to indicate what preferences typically support specific alternatives.

5. Discussion

In this paper we introduced ADDIS, a decision support system for evidence-based medicine. ADDIS was developed in the context of a scientific project aimed to enable better use of information technology in the transfer and analysis of clinical trials design and results. The long term vision was developed in collaboration with a steering group composed of experts from the pharmaceutical industry, academia and the regulatory environment. Short term plans were developed with our 'customer', a regulatory assessor who oversaw the development. The design was further informed by several (completed and ongoing) case studies, such as a study of the benefit-risk profiles of second generation anti-depressants. Although the software has been presented to and used by experts in the field, no formal validation or usability studies have been conducted so far.

We presented a unifying data model for aggregate trial results, which is at the core of ADDIS. The model enables semi-automated generation of evidence synthesis and benefit-risk models implemented in ADDIS. All these components together allow for re-usable, re-analyzable repositories of trials and analyses to be maintained and shared among users. The value of the unifying data model is not to model the domain in detail, but to provide a uniform basis for automated evidence synthesis and decision modeling. As such, specific decision support systems may use domain specific information to further assist the decision maker. ADDIS makes use of some domain knowledge to support its primary goal: to enable the direct and indirect assessment of the comparative benefits and risks of different drugs based on all available evidence from clinical trials. For example, Arms always have a Dosing characteristic, and studies have a fixed list of characteristics that are relevant for clinical trials comparing the efficacy and safety of

ACCEPTED MANUSCRIPT

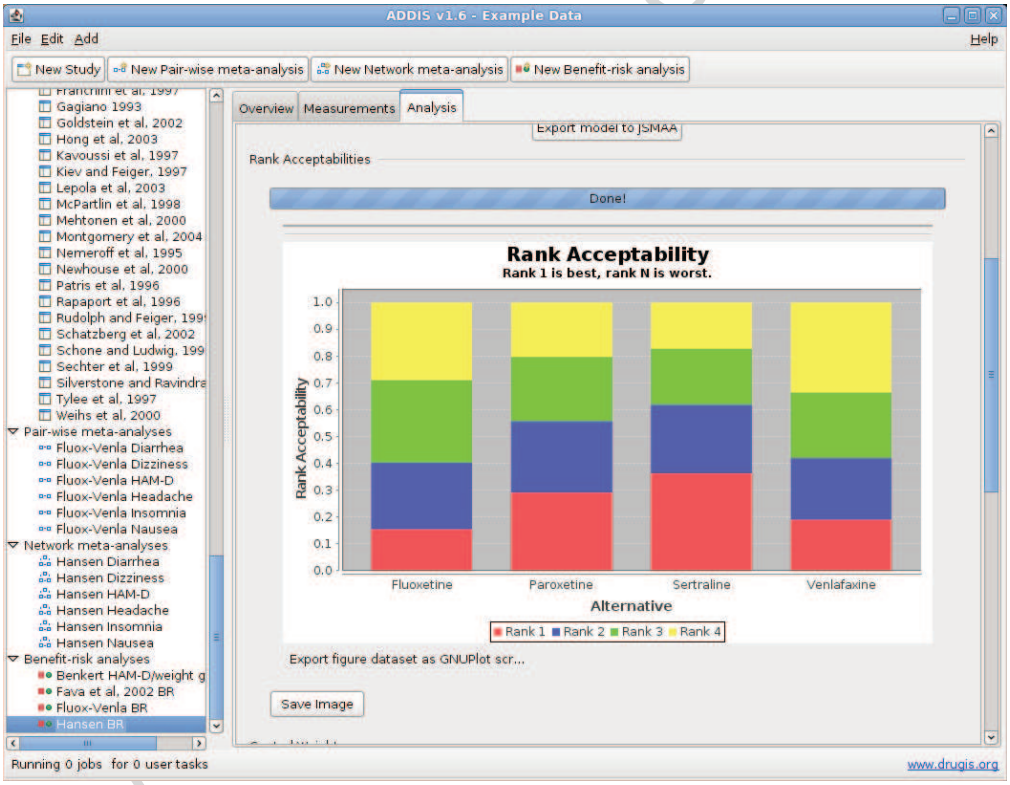


Figure 11: SMAA benefit-risk analysis results. The bars indicate the probability that each treatment is the best, second best, etc., given the preferences specified by the decision maker. In this case, the results indicate that there is a lot of uncertainty regarding which alternative is the best, but sertraline and paroxetine are somewhat more likely to be than venlafaxine and fluoxetine.

ACCEPTED MANUSCRIPT

drugs.

Multiple data models have been proposed for comprehensively storing information on the design and outcomes of clinical trials, e.g. the ClinicalTrial.gov DED and the CDISC standards. The minimal unifying data model implemented in ADDIS is not competing with these, but rather provides a target for conversion from them in order to enable semi-automated generation of evidence synthesis and decision models operating on the trial results. Traditionally the systematic reviewing process to perform a (network) meta-analysis takes a considerable amount of time and effort. While ADDIS does not address this problem directly, it does provide a uniform platform for analysis and data sharing that obviates the need for repeated data extraction.

To the best of our knowledge, ADDIS is the first system to implement decision models that are directly and explicitly based on the (synthesis of) clinical trials results. By making the involved trade-offs and the link between trial results and decision model recommendations visible, ADDIS can enable more transparent strategic health care decision making. ADDIS can also help in improving the reporting of systematic reviews since the included trials are represented explicitly, rather than only in data tables pre-processed for the purpose of evidence synthesis. The decisions made in mapping the data and applying the evidence synthesis models are thus clearly represented.

5.1. Limitations and future work

The decision modeling in ADDIS is based on the assumption that a structured database of relevant clinical trials is available. However, to acquire such a database is a difficult and time consuming. The initial phase of development has focused on drug regulation – a use case for which it is reasonable to assume that the data will be provided in whatever format requested. For other use cases, such as guideline formulation, this assumption is not justified. If the data is not available in a suitable format, a systematic review will have to be performed and the data input into ADDIS mostly manually, although the ClinicalTrials.gov import functionality can reduce the required work. However, once the input is done, the data is more valuable than the same set of trials extracted for e.g. Cochrane RevMan, as they can be reused for different types of analyses. To make ADDIS a useful tool for a wider audience, functionality that further increases the efficiency of systematic reviewing should be added, possibly by implementing automated information extraction methods.

Until now, approximately 100 clinical trials were entered for the case studies. To assess the usefulness of ADDIS in various medical domains more trials should be entered. However, as their input is mostly manual, this is an expensive and time-consuming process. Also, as the trial database gets larger, the study selection step for evidence synthesis can get cumbersome with the current implementation. More intelligent study matching/filtering (e.g. with the different characteristics) should be explored for lowering the user's work load. This may require explicit modeling of some of the aspects that are currently stored as plain text, such as the patient eligibility criteria.

The scope of the unifying data model could be extended to support other types of evidence synthesis, such as meta-regression and stratified analyses. These possible extensions may introduce covariates at different levels, e.g. the time at which an outcome measure was assessed, the dosage level for a treatment, the baseline severity of illness in an arm,

ACCEPTED MANUSCRIPT

the length of the placebo washout phase of a study, or within-arm correlation of two or more outcome measures. As such, it will be a challenge to introduce these rather complex distinctions without making the generation of (network) meta-analyses impossible.

ADDIS enables generation of benefit-risk decision models that use aggregate level, possibly synthesized, clinical trial data as part of their input. However, health care decisions can include evaluation dimensions not reported in clinical trials (e.g. convenience of administration or storage), which consequently can not be included in ADDIS. Also, economical decision models applied in health technology assessment often do take into account the primary clinical endpoints of interest with high quality evidence, but seldom include high-quality adverse event sources [69]. We acknowledge that adverse event reporting in general is inferior to clinical endpoint reporting due to various reasons. These include the rareness of some adverse events, the fact that most clinical trials are powered to show efficacy (which typically requires smaller sample size than detecting adverse events) and inconsistent reporting of adverse event data [83]. Decision models based on evidence synthesis can help improve the included evidence on adverse events, but it may be necessary to include other evidence sources to consider the rarest events. To consider these and other use cases, future research should address semi-automated generation of a wider range of decision models and their implementation in ADDIS.

Role of the funding source

This study was performed in the context of the Escher project (T6-202), a project of the Dutch Top Institute Pharma. The funding source had no direct involvement with the research presented in this paper.

References

[1] E. H. Shortliffe, B. G. Buchanan, A model of inexact reasoning in medicine, Math Biosci 23 (3–4) (1975) 351–379. doi:10.1016/0025-5564(75)90047-4.

[2] Evidence-Based Medicine Working Group, Evidence-based medicine. A new approach to teaching the practice of medicine, JAMA 268 (17) (1992) 2420–2425. doi:10.1001/jama.1992.03490170092032.

[3] A. J. Sutton, N. J. Cooper, D. R. Jones, Evidence synthesis as the key to more coherent and efficient research, BMC Med Res Methodol 9 (29) (2009) e–publication. doi:10.1186/1471-2288-9-29.

[4] J. Higgins, S. Green (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009], The Cochrane Collaboration, 2009, available from http://www.cochrane-handbook.org.

[5] I. Sim, D. K. Owens, P. W. Lavori, G. D. Rennels, Electronic trial banks: A complementary method for reporting randomized trials, Med Decis Making 20 (4) (2000) 440–450. doi:10.1177/0272989X0002000408.

[6] D. A. Zarin, N. C. Ide, T. Tse, W. R. Harlan, J. C. West, D. A. B. Lindberg, Issues in the registration of clinical trials, JAMA 297 (19) (2007) 2112–2120. doi:10.1001/jama.297.19.2112.

[7] I. Sim, A. W. Chan, A. M. Gulmezoglu, T. Evans, T. Pang, Clinical trial registration: transparency is the watchword, Lancet 367 (9523) (2006) 1631–1633. doi:10.1016/S0140-6736(06)68708-4.

[8] D. Ghersi, M. Clarke, J. Berlin, A. M. Guelmezoglu, R. Kush, P. Lumbiganon, D. Moher, F. Rockhold, I. Sim, E. Wager, Reporting the findings of clinical trials: a discussion paper, Bulletin of the World Health Organization 86 (6) (2008) 492–493. doi:10.2471/BLT.08.053769.

ACCEPTED MANUSCRIPT

[9] J. M. Grimshaw, N. Santesso, M. Cumpston, A. Mayhew, J. McGowan, Knowledge for knowledge translation: the role of the Cochrane Collaboration, J Contin Educ Health Prof 26 (1) (2006) 55–62. doi:10.1002/chp.51.

[10] A. J. J. Wood, Progress and deficiencies in the registration of clinical trials, N Engl J Med 360 (8) (2009) 824–830. doi:10.1056/NEJMsr0806582.

[11] T. Tervonen, G. van Valkenhoef, E. Buskens, H. L. Hillege, D. Postmus, A stochastic multi-criteria model for evidence-based decision making in drug benefit-risk analysis, Stat Med 30 (12) (2011) 1419–1428. doi:10.1002/sim.4194.

[12] P. M. Coplan, R. A. Noel, B. S. Levitan, J. Ferguson, F. Mussen, Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines, Clin Pharmacol Ther 89 (2) (2011) 312–315. doi:10.1038/clpt.2010.291.

[13] F. Mussen, S. Salek, S. Walker, A quantitative approach to benefit-risk assessment of medicines – part 1: The development of a new model using multi-criteria decision analysis, Pharmacoepidemiol Drug Saf 16 (Suppl. I) (2007) S12–S15. doi:10.1002/pds.1435.

[14] G. van Valkenhoef, T. Tervonen, J. Zhao, B. de Brock, H. L. Hillege, D. Postmus, Multi-criteria benefit-risk assessment using network meta-analysis, J Clin Epidemiol (in press). doi:10.1016/j.jclinepi.2011.09.005.

[15] S. Kiritchenko, B. de Bruijn, S. Carini, J. Martin, I. Sim, ExaCT: automatic extraction of clinical trial characteristics from journal publications, BMC Med Inform Decis Mak 10 (2010) 56. doi:10.1186/1472-6947-10-56.

[16] CDISC, CDISC 2004 research project on attitudes, adoption, and usage of data collection technologies and data interchange standards; executive summary (Sept 2005).

[17] K. El Emam, E. Jonker, M. Sampson, K. Krleza-Jeric, A. Neisa, The use of electronic data capture tools in clinical trials: Web-survey of 259 canadian trials, J Med Internet Res 11 (1) (2009) e8. doi:10.2196/jmir.1120.

[18] Tufts, CSDD and CDISC, Study on the adoption and attitudes of electronic clinical research technology solutions and standards; summary of results (2007).
URL http://www.cdisc.org/stuff/contentmgr/files/0/e35818eab8d8cc7b9d159ffeba5cdda5/misc/tuftstop30rdkjan08.pdf

[19] FDA, Guidance for industry: Providing regulatory submissions in electronic format drug establishment registration and drug listing, US Food and Drug Administration (FDA). OMB Control No. 0910-0045 (May 2009).

[20] CDISC and FDA, Walking down the critical path: The application of data standards to FDA submissions, a Discussion Paper by CDISC and FDA (February 2005).

[21] C. D. Mulrow, Rationale for systematic reviews, BMJ 309 (1994) 597–599.

[22] R. B. Haynes, K. A. McKibbon, N. L. Wilczynski, S. D. Walter, S. R. Werre, Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey, BMJ 330 (2005) 1179. doi:bmj.38446.498542.8Fv1.

[23] K. Dickersin, D. Rennie, Registering clinical trials, J Am Med Inform Assoc 290 (4) (2003) 516–523. doi:10.1001/jama.290.4.516.

[24] K. Lee, P. Bacchetti, I. Sim, Publication of clinical trials supporting successful new drug applications: A literature analysis, PLoS Med 5 (9) (2008) e191. doi:10.1371/journal.pmed.0050191.

[25] R. J. Simes, Publication bias - the case for an international registry of clinical-trials, J Clin Oncol 4 (10) (1986) 1529–1541.

[26] A. T. McCray, N. C. Ide, Design and implementation of a national clinical trials registry 7 (3) (2000) 313–323. doi:10.1136/jamia.2000.0070313.

[27] ICTRP. About trial registration: Organizations with policies [online] (2010). Archived at http://www.webcitation.org/5mZpckdBf [cited 6 January 2010].

[28] K. Krleza-Jeric, A.-W. Chan, K. Dickersin, I. Sim, J. Grimshaw, C. Gluud, the Ottawa Group, Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1), BMJ 330 (7497) (2005) 956–958. doi:10.1136/bmj.330.7497.956.

[29] J. Kaiser, Making clinical data widely available, Science 322 (5899) (2008) 217–218. doi:10.1126/science.322.5899.217.

[30] D. A. Zarin, T. Tse, Medicine - moving toward transparency of clinical trials, Science 319 (5868) (2008) 1340–1342. doi:10.1126/science.1153632.

ACCEPTED MANUSCRIPT

[31] A.-W. Chan, Bias, spin, and misreporting: Time for full access to trial protocols and results, PLoS Med 5 (11) (2008) e230. doi:10.1371/journal.pmed.0050230.

[32] I. Sim, C. G. Chute, H. Lehmann, R. Nagarajan, M. Nahm, R. H. Scheuermann, Keeping raw data in context, Science 323 (5915) (2009) 713a. doi:10.1126/science.323.5915.713a.

[33] FDA, US Food and Drug Administration Amendments Act (FDAAA), Section 801 (2007).

[34] ClinicalTrials.gov. Linking to clinicaltrials.gov [online] (April 2009). Archived at http://www.webcitation.org/5mZlpYkGW [cited 6 January 2010].

[35] European Medicines Agency. About EU clinical trials register [online] (2011). Archived at http://www.webcitation.org/5yNkGPZZX [cited 2 May 2011].

[36] D. M. Caldwell, N. J. Welton, A. E. Ades, Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency, J Clin Epidemiol 63 (8) (2010) 875 – 882. doi:10.1016/j.jclinepi.2009.08.025.

[37] P. K. Honig, Systematic reviews and meta-analyses in the new age of transparency, Clin Pharmacol Ther 88 (2) (2010) 155–158. doi:10.1038/clpt.2010.124.

[38] J. P. A. Ioannidis, Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses, Can Med Assoc J 181 (8) (2009) 488–493. doi:10.1503/cmaj.081086.

[39] K. Dickersin, E. Manheimer, S. Wieland, K. A. Robinson, C. Lefebvre, S. McDonald, Development of the Cochrane Collaboration's central register of controlled clinical trials, Eval Health Prof 25 (38) (2002) 38–64. doi:10.1177/016327870202500104.

[40] M. Starr, I. Chalmers, The evolution of The Cochrane Library, 1988–2003 (2003) [cited 21 October 2008].
URL http://www.update-software.com/history/clibhist.htm.

[41] A. R. Jadad, D. J. Cook, A. Jones, T. P. Klassen, P. Tugwell, M. Moher, D. Moher, Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals, JAMA 280 (1998) 278–280. doi:10.1001/jama.280.3.278.

[42] Biomedical Research Integrated Domain Group (BRIDG), BRIDG Model Release 3.0.3 Users Guide (2010).
URL http://gforge.nci.nih.gov/frs/?group_id=342

[43] ClinicalTrials.gov. Clinicaltrials.gov protocol data element definitions (draft) [online] (November 2009). Archived at http://www.webcitation.org/5mYe6OxvP [cited 5 January 2010].

[44] ClinicalTrials.gov. Clinicaltrials.gov "basic results" data element definitions (draft) [online] (September 2009) [cited 6 January 2010].

[45] S. W. Tu, M. Peleg, S. Carini, M. Bobak, J. Ross, D. Rubin, I. Sim, A practical method for transforming free-text eligibility criteria into computable criteria, J Biomed Inform 44 (2) (2011) 239–250. doi:10.1016/j.jbi.2010.09.007.

[46] I. Sim, S. Carini, S. Tu, R. Wynden, P. BH, S. Mollah, D. Gabriel, H. Hagler, R. Scheuermann, H. Lehmann, K. Wittkowski, M. Nahm, S. Bakken, The human studies database project: Federating human studies design data using the ontology of clinical research, in: Proceedings of the AMIA CRI Summit 2010, 2010.

[47] S. Carini, B. H. Pollock, H. P. Lehmann, S. Bakken, E. M. Barbour, D. Gabriel, H. K. Hagler, C. R. Harper, S. A. Mollah, M. Nahm, H. H. Nguyen, R. H. Scheuermann, I. Sim, Development and evaluation of a study design typology for human research, AMIA Annu Symp Proc 2009 (2009) 81–85.

[48] Y. M. Kong, C. Dahlke, Q. Xiang, Y. Qian, D. Karp, R. H. Scheuermann, Toward an ontology-based framework for clinical research databases, J Biomed Inform 44 (1) (2011) 48–58. doi:10.1016/j.jbi.2010.05.001.

[49] R. H. Scheuermann. Ontology-based extensible data model [online] (2010). Archived at http://www.webcitation.org/5y3tihqvB [cited 19 April 2011].

[50] J. J. Cimino, Review paper: coding systems in health care, Methods Inf Med 35 (1996) 273–284.

[51] S. Nelson, M. Schopen, A. Savage, J.-L. Schulman, N. Arluk, The MeSH translation maintenance system: Structure, interface design, and implementation, in: Proceedings of the 11th World Congress on Medical Informatics, San Francisco, 2004, pp. 67–69.

[52] S. Schulz, S. Hanser, U. Hahn, J. Rogers, The semantics of procedures and diseases in SNOMED CT, Methods Inf Med 45 (2006) 354–358.

ACCEPTED MANUSCRIPT

[53] The International Health Terminology Standards Development Organisation. SNOMED CT components [online] (2011). Archived at http://www.webcitation.org/5yNmLpPkv [cited 2 May 2011].

[54] D. A. B. Lindberg, B. L. Humphreys, A. T. McCray, The unified medical language system, Methods Inf Med 32 (1993) 281–291.

[55] O. Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, Nucleic Acids Res 32 (2004) D267–270. doi:10.1093/nar/gkh061.

[56] A. M. Cohen, W. R. Hersh, A survey of current work in biomedical text mining, Brief Bioinform 6 (2005) 57–71. doi:10.1093/bib/6.1.57.

[57] S. M. Meystre, G. K. Savova, K. C. Kipper-Schuler, J. F. Hurdle, Extracting information from textual documents in the electronic health record: a review of recent research, Yearb Med Inform (2008) 128–144.

[58] S.-L. T. Normand, Meta-analysis: formulating, evaluating, combining, and reporting, Stat Med 18 (3) (1999) 321–359. doi:10.1002/(SICI)1097-0258(19990215)18:3<321::AID-SIM28>3.0.CO;2-P.

[59] T. Lumley, Network meta-analysis for indirect treatment comparisons, Stat Med 21 (16) (2002) 2313–2324. doi:10.1002/sim.1201.

[60] G. Lu, A. E. Ades, Combination of direct and indirect evidence in mixed treatment comparisons, Stat Med 23 (20) (2004) 3105–3124. doi:10.1002/sim.1875.

[61] G. Salanti, J. P. T. Higgins, A. E. Ades, J. P. A. Ioannidis, Evaluation of networks of randomized trials, Stat Methods Med Res 17 (3) (2008) 279–301. doi:10.1177/0962280207080643.

[62] G. Lu, A. E. Ades, Assessing evidence inconsistency in mixed treatment comparisons, J Am Stat Assoc 101 (474) (2006) 447–459. doi:10.1198/016214505000001302.

[63] S. Dias, N. J. Welton, D. M. Caldwell, A. E. Ades, Checking consistency in mixed treatment comparison meta-analysis, Stat Med 29 (7-8, Sp. Iss. SI) (2010) 932–944. doi:10.1002/sim.3767.

[64] D. M. Caldwell, A. E. Ades, J. P. T. Higgins, Simultaneous comparison of multiple treatments: combining direct and indirect evidence, BMJ 331 (7521) (2005) 897–900. doi:10.1136/bmj.331.7521.897.

[65] A. Cipriani, T. A. Furukawa, G. Salanti, J. R. Geddes, J. P. T. Higgins, R. Churchill, N. Watanabe, A. Nakagawa, I. M. Omori, H. McGuire, M. Tansella, C. Barbui, Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis, Lancet 373 (9665) (2009) 746 – 758. doi:10.1016/S0140-6736(09)60046-5.

[66] C. Stettler, S. Allemann, S. Wandel, A. Kastrati, M. C. Morice, A. Schoemig, M. E. Pfisterer, G. W. Stone, M. B. Leon, J. Suarez de Lezo, J.-J. Goy, S.-J. Park, M. Sabate, M. J. Suttorp, H. Kelbaek, C. Spaulding, M. Menichelli, P. Vermeersch, M. T. Dirksen, P. Cervinka, M. De Carlo, A. Erglis, T. Chechi, P. Ortolani, M. J. Schalij, P. Diem, B. Meier, S. Windecker, P. Juni, Drug eluting and bare metal stents in people with and without diabetes: collaborative network meta-analysis, BMJ 337 (2008) a1331. doi:10.1136/bmj.a1331.

[67] A. J. Sutton, J. P. T. Higgins, Recent developments in meta-analysis, Stat Med 27 (5) (2008) 625–650. doi:10.1002/sim.2934.

[68] R. B. Haynes, P. J. Devereaux, G. H. Guyatt, Clinical expertise in the era of evidence-based medicine and patient choice, Evid Based Med 7 (2002) 36–38. doi:10.1136/ebm.7.2.36.

[69] N. Cooper, D. Coyle, K. Abrams, M. Mugford, A. Sutton, Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997, J Health Serv Res Pol 10 (4) (2005) 245–250. doi:10.1258/135581905774414187.

[70] J. J. Guo, S. Pandey, J. Doyle, B. Bian, Y. Lis, D. W. Raisch, A Review of Quantitative Risk-Benefit Methodologies for Assessing Drug Safety and Efficacy-Report of the ISPOR Risk-Benefit Management Working Group, Value Health 13 (5) (2010) 657–666. doi:10.1111/j.1524-4733.2010.00725.x.

[71] D. Ouellet, Benefit-risk assessment: the use of clinical utility index, Expert Opin Drug Saf 9 (2) (2010) 289–300. doi:10.1517/14740330903499265.

[72] B. Roy, Multicriteria Methodology for Decision Analysis, Kluwer Academic Publishers, Dordrecht, 1996.

[73] R. Lahdelma, J. Hokkanen, P. Salminen, SMAA - stochastic multiobjective acceptability analysis, Eur J Oper Res 106 (1) (1998) 137–143. doi:10.1016/S0377-2217(97)00163-X.

[74] J. C. Felli, R. A. Noel, P. A. Cavazzoni, A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives, Med Decis Making 29 (1) (2009) 104–115. doi:10.1177/0272989X08323299.

ACCEPTED MANUSCRIPT

[75] L. D. Lynd, B. J. O'Brien, Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis, J Clin Epidemiol 57 (8) (2004) 795–803. doi:10.1016/j.jclinepi.2003.12.012.

[76] M. Egger, G. D. Smith, A. N. Phillips, Meta-analysis: principles and procedures, BMJ 315 (7121) (1997) 1533–1537.

[77] R. Lahdelma, P. Salminen, SMAA-2: Stochastic multicriteria acceptability analysis for group decision making, Oper Res 49 (3) (2001) 444–454. doi:10.1287/opre.49.3.444.11220.

[78] T. Tervonen, J. R. Figueira, A survey on stochastic multicriteria acceptability analysis methods, Journal of Multi-Criteria Decision Analysis 15 (1–2) (2008) 1–14. doi:10.1002/mcda.407.

[79] G. van Valkenhoef, T. Tervonen, B. de Brock, H. Hillege, Algorithmic parametrization of mixed treatment comparisons, Stat Comput (in press). doi:10.1007/s11222-011-9281-9.

[80] T. Tervonen, JSMAA: an open source software for SMAA computations, in: C. Henggeler Antunes, D. Rios Insua, L. Dias (Eds.), Proceedings of the 25th Mini-EURO conference on Uncertainty and Robustness in Planning and Decision Making (URPDM2010), Coimbra, Portugal, 2010.
URL http://drugis.org/files/tervonen-urpdm2010.pdf

[81] S. Lewis, M. Clarke, Forest plots: trying to see the wood and the trees, BMJ 322 (2001) 1479–1480. doi:10.1136/bmj.322.7300.1479.

[82] T. Tervonen, R. Lahdelma, Implementing stochastic multicriteria acceptability analysis, Eur J Oper Res 178 (2) (2007) 500–513. doi:10.1016/j.ejor.2005.12.037.

[83] H.-G. Eichler, F. Pignatti, B. Flamion, H. Leufkens, A. Breckenridge, Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma, Nat Rev Drug Discov 7 (10) (2008) 818–836. doi:10.1038/nrd2664.

ACCEPTED MANUSCRIPT