NOR-DMARD data management: implementation of data capture from electronic health records

I.C. Olsen, E.A. Haavardsholm, E. Moholt, T.K. Kvien, E. Lie

Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway.

Inge Christoffer Olsen, PhD
Espen André Haavardsholm, MD, PhD
Ellen Moholt, BSc
Tore Kristian Kvien, Prof. of Medicine
Elisabeth Lie, MD, PhD

Please address correspondence to:
Inge Christoffer Olsen, PhD,
Department of Rheumatology,
Diakonhjemmet Hospital,
P.O. Box 23 Vinderen,
N-0319 Oslo, Norway.
E-mail: inge.christoffer.olsen@gmail.com

ABSTRACT

Objective. *The use of electronic health records (EHR) is an essential part of modern health care, and electronic data capture (EDC) has become essential for managing clinical trials. Usually, these two entities are independent of each other, and transfer from one system to another is done manually. Our aim was to develop a method to capture data directly from the EHR system and transfer them into an EDC system for the NORwegian Disease-Modifying Anti-Rheumatic Drugs (NOR-DMARD) registry.*

Methods. *All rheumatology departments contributing to NOR-DMARD had implemented a structured EHR system. Data are extracted locally and securely transferred to the study data management once a month. The study data management then parse the data into a readable format for the EDC and import the data. Once the data is in the EDC, they are available to all authorized researchers and downloadable in a preferred format.*

Results. *From May 2012 to August 2014 almost 6400 visits in 3400 patients treated with biologics have been successfully registered in the EDC system. Previously, NOR-DMARD used standard paper-based case report forms (CRFs), with a substantial cost for data entry. Setting up and maintaining the EDC system required some investments, but the amount saved from avoiding paper handling has made the shift into EDC profitable. In addition to this, gains have been made in administration and data quality.*

Conclusion. *The transition from paper and pencil format to a fully electronic data management system in NOR-DMARD has had obvious advantages regarding feasibility, cost, data quality and accessibility of the data.*

Introduction

The NORwegian Disease Modifying Anti-Rheumatic Drugs (NOR-DMARD) registry has since 2000 been recording disease activity (including RA core set measures), quality of life measures and adverse events during DMARD treatment in patients with inflammatory joint diseases in five different Norwegian rheumatology departments (1). Initially, any adult patient with inflammatory arthropathies (rheumatoid arthritis (RA), psoriatic arthritis (PsA), ankylosing spondylitis (AS), spondyloartritis (SpA), adult juvenile idiopathic arthritis (JIA), undifferentiated arthritis) starting a new treatment with any synthetic or biologic DMARD treatment was included in the registry. In 2011 the steering committee decided to focus on biologic DMARD treatment only, and a new protocol was written and then implemented from 2012 after approval by the Regional Ethics Committee South East (ClinicalTrials.gov identifier NCT01581294). There have been changes to the data collection and visit schedule throughout the study since 2001; the current assessments and visit schedule are presented in Table I.

Methods

Data management principles
Initially, the data capture was based on paper case report forms (CRFs). Study nurses, physicians and patients would fill in the appropriate sections of the CRF. The original CRF pages were mailed to a Contract Research Organization (CRO) to be entered into a Microsoft Access database, while a copy of the CRF was kept at the centre. Filing and storage of the paper-CRFs as well as query management were handled by the CRO. By 2011 the amount of original CRFs was extensive, amounting to 560 shelf-metres. In addition, there were issues with data entry errors and that the data entry process was expensive and time consuming. With the change in protocol, the steering committee decided to move from paper-based to electronic data

Table I. Visiting schedule.

Visits	1	2	3	4	5	6	7, etc.
Months	0	3	6	9	12	18	every 6 months
Assessments by research nurse/physician							
Assessment of disease activity (VAS)	x	x	x	x	x	x	x
Swollen joints (28)	x	x	x	x	x	x	x
Tender joints (28)	x	x	x	x	x	x	x
Swelling of feet and ankles	x	x	x	x	x	x	x
Tenderness of feet and ankles	x	x	x	x	x	x	x
Medication (bDMARDs, sDMARDs, corticosteroids)	x	x	x	x	x	x	x
Intraarticular corticosteroid injections	x	x	x	x	x	x	x
AEs	x	x	x	x	x	x	x
Cross-check of patient reported comorbidity	x	x	x	x	x	x	x
Patient assessments							
Smoking	x						
Years of education	x						
Pain/fatigue/global disease activity (VAS)	x	x	x	x	x	x	x
MHAQ	x	x	x	x	x	x	x
EQ-5D	x	x	x	x	x	x	x
PASS/MCII	x	x	x	x	x	x	x
BASDAI	x	x	x	x	x	x	x
RAID score	x	x	x	x	x	x	x
Work (WPAI)	x	x	x	x	x	x	x
Employment	x	x	x	x	x	x	x
Comorbidity (patient-reported)	x	x	x	x	x	x	x
Biochemical assessments							
ACPA/IgM-RF/HLA-B27	x						
ESR	x	x	x	x	x	x	x
CRP	x	x	x	x	x	x	x
Biobank (DNA, serum, plasma)	x	x					
Radiological assessments							
Conventional radiographs of hands and feet	x				x		x

VAS: Visual Analogue Scale; MHAQ: Modified Health Assessment Questionnaire; EQ-5D: EuroQoL 5 Dimensions; PASS: Patient Acceptable Symptom State; MCII: Minimal Clinically Important Improvement; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; RAID: Rheumatoid Arthritis Impact of Disease; WPAI: Work Productivity and Activity Impairment Questionnaire; ESR: Erythrocyte Sedimentation Rate; CRP: C-Reactive Protein; ACPA: Anti-Citrullinated Protein Antibody; RF: Rheumatoid Factor.

capture in order to avoid further paper handling.

All rheumatology departments contributing to NOR-DMARD had by 2011 implemented a structured electronic health record (EHR) system, GoTreatIT™ (DiaGraphIT AS, Kristiansand, Norway). This system enhances disease monitoring, for example according to the treat-to-target principles,(2) by providing a graphic as well as numeric display of the longitudinal data of measures that have been performed (Fig. 1a-b). The system also facilitates a study tool where assessments done during clinic visits can be arranged into a study structure and with a tailored data extraction according to the study CRF. We

(the study management team) wanted to take advantage of this tool to capture the NOR-DMARD data. However, there were some limitations to this solution that required attention:

1. The study tool was quite rigid and limited to pre-specified modules;
2. Some protocol-specific information was not possible to capture adequately (*e.g.* adverse events);
3. There was no functionality for audit trailing or query handling;
4. The data were stored locally so a system had to be set up to merge the locally extracted de-identified data into a central database.

In order to meet these limitations, we set up an Electronic Data Capture

(EDC) system (Viedoc™, Pharma Consulting Group, Uppsala, Sweden) to capture study data using electronic case report forms (e-CRFs). This is a generic and flexible web-based system, compliant with all relevant regulations in North America, Europe and Japan (including 21 CFR Part 11, CSUCI, ICH GCP, HIPAA, PuL and EU Annex 11). We used this system as a central database.

The workflow as it is currently being performed is described below:

1. The patient is included into the EDC system with patient information such as patient initials, date of birth, current diagnosis and initiated biologic treatment, in addition to some information not collected in the EHR.
2. The EDC system generates a unique patient number, which is then registered in the EHR. This key enables the transfer from one system to the other.
3. Study data are then registered in the EHR, both by the treating physician/nurse (*e.g.* joint counts, acute phase reactants) and the patient (patients reported outcomes).
4. Some additional data (*e.g.*, adverse events) are registered directly to the EDC system.

Some composite indices and total scores, such as the disease activity score-28 (DAS28), the modified health assessment questionnaire (MHAQ) score and the ankylosing spondylitis disease activity score (ASDAS), are calculated in the EHR and/or EDC systems, while additional syntaxes for commonly used response criteria and EQ-5D are run on data exported from the EDC.

Once a month each centre extracts de-identified data in an XML-format from the centre's EHR system. The contents of the extracted data are pre-specified and in accordance with the study protocol and e-CRF. The XML-file is then uploaded to a secure location accessible to the study data management, which uses a SAS program to parse the XML-files to a flat file format readable in the EDC system. The SAS program also runs a validation check against patient numbers, dates of birth and visit dates. The flat files are subsequently imported into the EDC system using
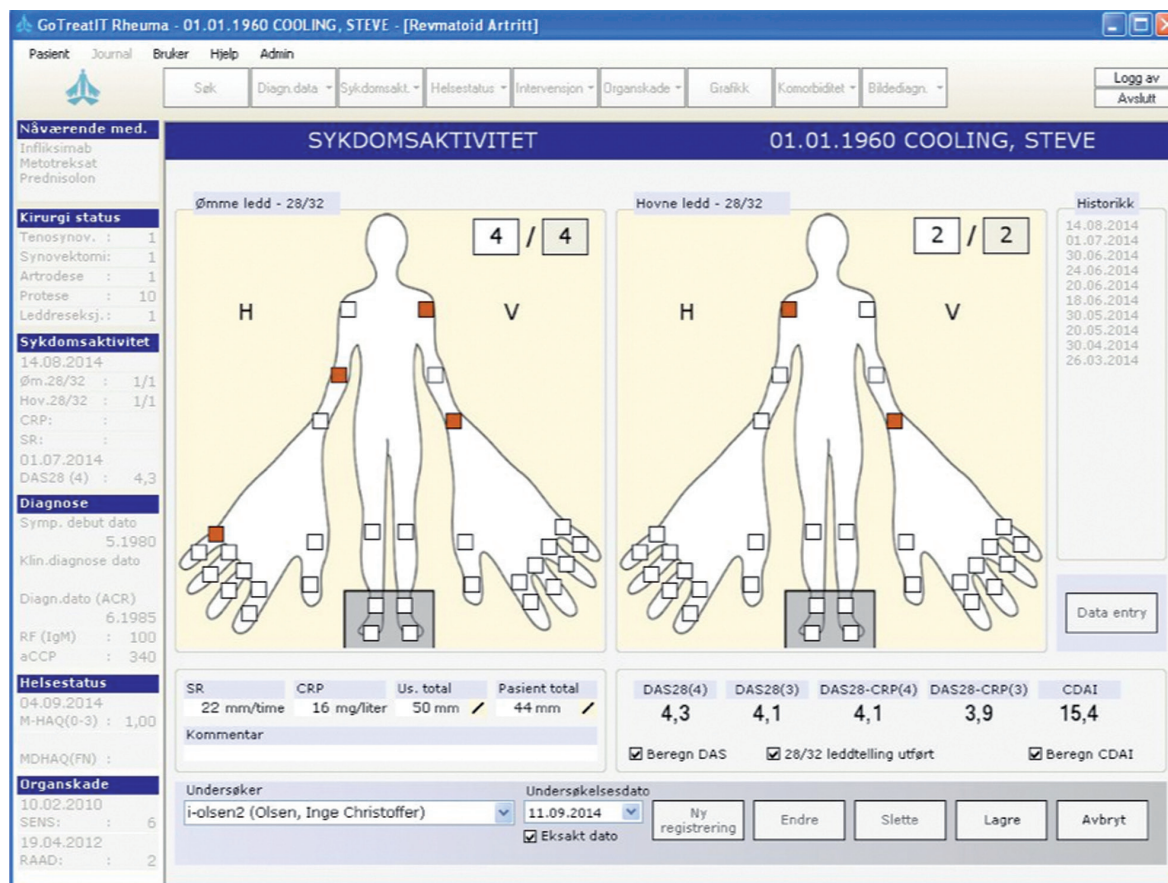
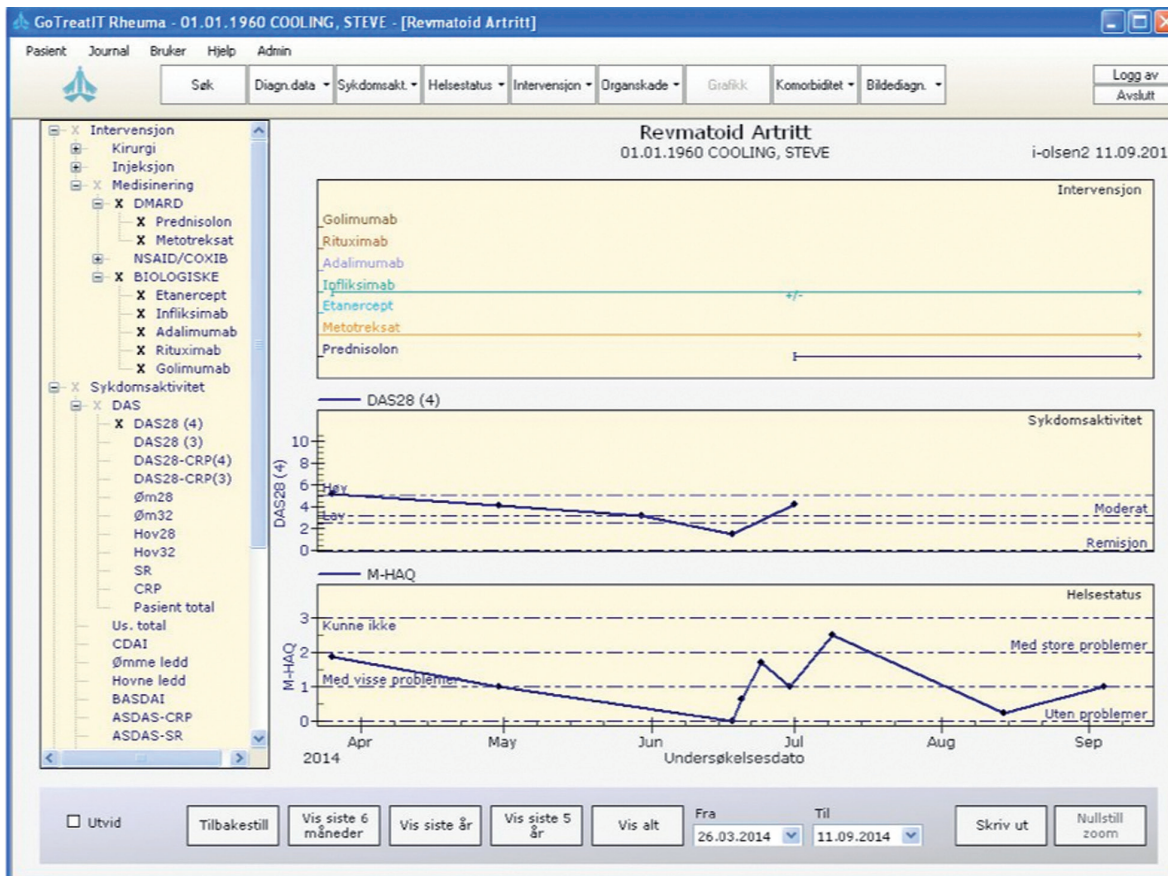**Fig. 1a.**
Data capture in GoTreatIT™

Fig. 1b.
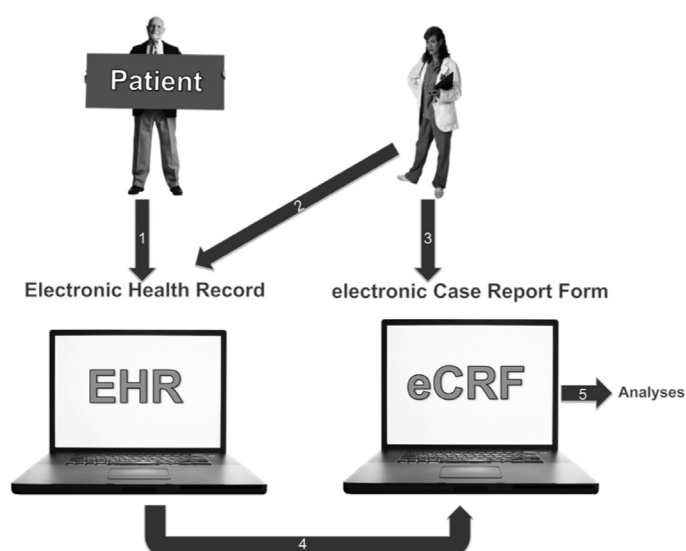Longitudinal display of disease measures and treatment in GoTreatIT™

1. Patient registered outcomes (PROs) such as M-HAQ and RAID are recorded into the EHR system by the patient;

2. The treating nurse/physician record clinical information such as joint counts and acute phase reactants into the EHR system;

3. Any information not captured in an easily retrievably way in the EHR is registered directly in the eCRF (*e.g.* adverse events, infusions);

4. Once a month each centre extracts de-identified data in an XML format from the centre's EHR system and uploads the file to a secure web location. The XML-files are then parsed to a format readable by the eCRF system using SAS, and then uploaded into the eCRF;

5. Data in the eCRF is available for analyses at any time in SPSS-, SAS-, or flat file format.

Fig. 2. Data flow in the NOR-DMARD registry.

the system's import routine, merging by patient number. The system facilitates an audit trail, and all changes made to data records are registered and tracked. A patient's e-CRF with all collected data is then accessible for both site and study data management. Any queries arising from the import is recorded in the EDC system, and resolved by the corresponding centre study personnel. There is no paper involved in this process. The data flow is illustrated in Figure 2.

The main study database embedded in the EDC system is available for all authorized researches and downloadable in the preferred format, *e.g.* ASCII, SAS or SPSS. Any database errors discovered during the import procedure or subsequent data analyses are easily corrected in the EDC system, either directly or through a query to the appropriate site. In the previous system the error correction procedure was much more elaborate, relying on paper queries and manual changes in the analysis data file. Similarly to the system in countries such as Sweden and Denmark, every citizen in Norway has a national ID-number, an 11-digit personal identifier that is used in all public registries as well as in the major national health registries, such as the Cancer Registry, the Cause of Death Registry, the Norwegian Patient Registry and the Norwegian Prescription Database. This personal ID-number is not part of the NOR-DMARD database,

which is de-identified, but is stored as source data in the EHR system (GoTreatIT) at each participating centre. The NOR-DMARD study protocol and the patient consent open for linkage to other registers by the personal ID-number to obtain for example validated (serious) adverse event and comorbidity data.

Results
From May 2012 to August 2014, almost 6400 visits in 3400 patients on biologic treatment have been successfully included into the EDC system. A Contract Research Organization (CRO) handled the previous paper-CRF system at a price of 14 EUR per visit/CRF. Handling this amount in the previous paper-based system would have a total cost of 88.000 EUR. The initial set-up costs (including licensing fees) of the EDC system amounted to 18.000 EUR, and the yearly licensing fees are about 1.800 EUR giving a total cost by August 2014 of about 24.000 EUR. These numbers exclude the costs of the EHR system, as this was implemented on all study sites independent of the study. Some internal data management costs are not included here, but the money saved has already made the shift into EDC profitable. In addition to this, the paper CRF system required considerable manual resources at each participating centre, including entering, copying and mailing of CRFs and administration of inclusion IDs.

Discussion
The transition from paper and pencil format to a fully electronic data management system in NOR-DMARD has had obvious advantages regarding feasibility, cost and also data quality. In addition, data for analyses can be extracted within minutes from the central database, with updates from the centres every month, compared to previously biannual updates of the database through manual routines.

However, there are also some problems and challenges with the new methodology. The export/import routine is complex and relies on SAS programming expertise. As of today there is only one person within the study management with the necessary knowledge to perform the import into the EDC system. This makes the system vulnerable. The export/import routine is quite time consuming, currently approximately 10 hours per transfer. However, compared to manual entering of paper CRFs this is very efficient. There are also some problems with missing data in the EHR, especially regarding fulfillment of classification criteria since there is no alert to the study personnel to fill out this information when registering a patient. Some additional effort to make the study database as complete as possible is needed in the near future. Missing patient-reported data does occur, though not very frequently. We have adopted the position that we do not in-

terfere with patient registered data. In the entry system, patients cannot make entry errors (*e.g.* tick two boxes when only one should be ticked). They can avoid entering anything, but they are then prompted if they are certain of proceeding.

In conclusion, we have developed a novel methodology where we gather information from electronic health records in five different centres into a central trial database. We avoid time-consuming handling of paper-CRFs and have limited the occurrence of data entry errors since we import source data directly into the study database. Compared to a standard e-CRF solution we gain efficacy because we avoid redundant registration.

References
1. KVIEN TK, HEIBERG, LIE E *et al.*: A Norwegian DMARD register: prescriptions of DMARDs and biological agents to patients with inflammatory rheumatic diseases. *Clin Exp Rheumatol* 2005; 23 (Suppl. 39): S188-S94.
2. SMOLEN JS, ALETAHA D, BIJLSMA JWJ *et al.*: Treating rheumatoid arthritis to target: recommendations of an international task force. *Ann Rheum Dis* 2010; 69: 631-7.