Using big data from real-world Australian rheumatology encounters to enhance clinical care and research

G.O. Littlejohn[1], K.E. Tymms[2], T. Smith[3], H.T. Griffiths[4]

[1]Monash Rheumatology and Monash University, Clayton, VIC, Australia; [2]Canberra Rheumatology, Canberra, ACT, Australia; [3]OPAL Rheumatology Ltd, Sydney, NSW, Australia; [4]Barwon Rheumatology Service, Geelong, VIC, Australia.

Abstract

Objective

OPAL (Optimising Patient outcomes in Australian rheumatoLogy) Rheumatology is an independent not for profit Australian clinical research organisation which is the custodian of one of the largest datasets of patients with rheumatic diseases in the world, containing real-world clinical data from more than >175,000 unique patients collected over more than 900,000 clinical consultations. We describe the evolution and outcomes of the OPAL dataset, with particular reference to the use of big data derived from real-world clinical encounters to enhance clinical care and research.

Methods

De-identified data are regularly extracted and aggregated from the electronic medical records (EMR) of consenting patients treated by approximately 100 rheumatologists around Australia. The EMR shared by OPAL clinicians was specifically customised for rheumatology and collects comprehensive information on demographics, disease history, activity and severity, co-morbidities, pathology, and medication use. In addition, OPAL captures multifaceted outcomes data from the patient perspective through a novel electronic patient-reported outcome (ePRO) delivery system which allows for health-related quality of life measures to be matched with clinical indices.

Results

Since inception in 2009, OPAL has produced 35 publications and abstracts. OPAL also provides real-world data to determine drug utilisation, efficacy and safety, elucidate the natural history of disease, highlight areas of unmet need, guide medical affairs and commercial strategy, and to support regulatory and reimbursement submissions.

Conclusion

The extensive, evolving and organic OPAL dataset reflects the complexities of clinical rheumatological practice. It provides unique opportunities to enhance clinical care and research.

Key words

OPAL, data set, rheumatology, Australia

Geoffrey O. Littlejohn, MD, FRACP
Kathleen E. Tymms, FRACP
Tegan Smith, PhD
Hedley T. Griffiths, FRACP

Please address correspondence to:
Geoffrey Littlejohn,
Suite H, Private Consulting Rooms,
Monash Medical Centre,
246 Clayton Road,
3168 Clayton, Victoria, Australia.
E-mail: geoff.littlejohn@monash.edu

Received on June 28, 2019; accepted in revised form on September 30, 2019.

Reprints will not be available from the author.

Funding: Roche Australia provided an unrestricted grant to initiate this project

Competing interests: none declared.

Introduction

OPAL (Optimising Patient outcomes in Australian rheumatoLogy) was established in 2009 as a Quality Use of Medicines Initiative with the support of an educational grant from Roche Pharmaceuticals Australia. There were two broad aims. The first was to facilitate the use of a common customised electronic medical record (EMR) by rheumatologists, in real-world clinical practice to enhance clinical care. The second was to collect data to answer questions relevant to contemporary clinical rheumatological practice. These aims are incorporated in OPAL's mission statement which is "to optimise patient related outcomes in rheumatic disease through audit and collaborative research obtained at point of care, and to disseminate to (and for the good of) the wider community".

Ten years later there are over 100 OPAL members practicing in 42 clinics representing one third of practising rheumatologists in Australia. Over 90 percent are in private practice reflecting the fact that the majority of patients with rheumatic diseases in the health-care system in Australia are managed in non-public hospital clinics. There are over >173,000 patients in the dataset across all rheumatological conditions, including for example, rheumatoid arthritis (RA) (42,000 patients), psoriatic arthritis (PsA) (12,053 patients), ankylosing spondylitis (AS) (4,279 patients) and systemic lupus erythematosus (SLE) (4,091 patients). Conditions seen uncommonly by any one rheumatologist can be identified in sufficient numbers that allow for detailed study when the whole dataset is examined. For instance, there are 1,235 patients with giant-cell arteritis. With data from over 900,000 clinical consultations spanning 10 years, OPAL is likely to be one of the largest collections of real-world rheumatology patient data. An additional 7,000 new patients are added to the dataset each quarter.

Because OPAL is completely observational and makes secondary use of clinical data, it is not a classic registry. OPAL rheumatologists agree to share their patient clinical EMR for data gathering. Hence, as new data are entered electronically into the patient EMR by their clinician, a process occurring with multiple clinician-patient consultations numerous times each day, the OPAL data are best described as a dynamic, evolving and organic dataset.

The OPAL project therefore involves extensive use of big data with considerable scope for rapid assessment of real-world management of a wide range of both common and rare rheumatic diseases. The purpose of this manuscript is to detail the evolution of the OPAL project and examine how this big dataset can be best used to enhance clinical care of patients with rheumatic diseases.

Materials and methods

To achieve the aims of OPAL a steering committee directed by volunteer clinical rheumatologists was established in 2009. Other clinicians were invited to participate as members of OPAL. An established EMR, named Audit4 ®(Software4Specialists, (S4S)) was identified and made available to all members. This software was specifically tailored to collect and collate information relevant to the rheumatological clinical consultation. Different workbooks, acting as in-programme apps, have been developed to allow easy collection, at the point-of-care, of relevant clinical findings, such as tender and swollen joint counts (TJC, SJC) with patient and physician global assessments (PtGA, PGA). Each workbook relates to a clinician designated (ie non-criteria) diagnosis to which the relevant ICD-10 diagnostic code is applied for subsequent auditing, for example rheumatoid arthritis. Routine pathology results are downloaded to the EMR allowing composite outcome scores, such as Disease Activity Score 28 with CRP or ESR (DAS28-CRP, DAS28-ESR), to be automatically calculated in real time. Electronic patient-reported outcome (ePRO) data are collected through secure email or iPAD tablet in the waiting room prior to clinical review with completed PRO scores appearing in the patients' EMR for immediate review. This adds qualitative self-report data on levels of pain, fatigue, physical function impairment and general well-being, and allows for calculation of scores

including Clinical Disease Activity Index (CDAI), Simple Disease Activity Index (SDAI), Routine Assessment of Patient Index Data 3 (RAPID 3), Health Assessment Questionnaire (HAQ) and Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue). The data contained in these real-world clinical encounters are owned by the individual clinician and is stored on the clinician's server. However, the data can be de-identified while on the server and uploaded to a central secure server and aggregated with the wider group, making it available for group interrogation. This is done every 3 months. This aggregated dataset which is de-identified to patient, doctor and clinic is the OPAL dataset and is owned and managed by the OPAL board of directors. Clinicians contributing clinical data to OPAL are not paid for their contributions. The activities of OPAL Rheumatology have received overarching ethics approval from the University of New South Wales Human Research Ethics Committee (HREC) based on a patient opt out consent arrangement and each study protocol is also approved by an independent HREC prior to analysis by the OPAL approved statistician. Written notification of this arrangement is placed in an obvious place in every clinician's office.

Figure 1 summarises the methodology used by OPAL.

In 2015, OPAL became independent of Roche (and the then second provider of an educational grant - Celgene), and became a not-for-profit company, called OPAL Rheumatology Ltd (OPAL). The basic aims of OPAL were unchanged. The first remains to continue to enhance clinical care, and the second research focus comprises two core elements, these being Key Research Questions (KRQs), and a medication-focussed audit, named OPAL BI.

The KRQs are required to address a valid contemporary clinical rheumatology question and are formalised and approved by the OPAL scientific committee and then presented to an independent HREC for approval. Some of these studies are done in collaboration with a sponsor, who may propose a relevant study question; the project however is managed by OPAL and the data are analysed, interpreted and published independently by OPAL.

The second core research element is the creation of a large medication orientated dataset filtered into visually accessible components by powerful and sophisticated data visualisation software which has advanced significantly in recent years (Tableau® (Tableau Software)). This information can show different interacting data such as diagnosis, patient demographics, reason for starting and stopping a drug, switching patterns, disease activity, line of therapy, persistence on drug and market share of a particular drug. This provides extensive contemporaneous real-world clinical information that is relevant to Health Care funding agencies (Government, Insurance Companies), Pharmaceutical Companies and other involved parties. This information is also available to individual OPAL members to allow for benchmarking and reflection of an individual's clinical outcomes compared to the total group. The OPAL BI dataset contains 37 million rows of data in the current iteration.

Figure 2 illustrates examples from the OPAL BI dataset using Tableau Software.

Results

Routine clinical data collection

Clinical observations during the patients' visits with an OPAL member are documented in the EMR. The parameters that are recorded include: patient demographics (sex and age), co-morbidities, the date of disease onset and the date that the patient was first and last seen. Medications for the rheumatological condition being managed are recorded, together with start and stop dates, as well as the reasons for initiating and ceasing a drug. A homunculus allows for rapid documentation of swollen and tender joints and facilitates a quick comparison with results from past consultations. Pathology and imaging reports are electronically transferred from providers and are incorporated into the patient's medical record. The EMR collates all clinical information for each patient and automatically calculates disease activity scores (for example, DAS28-CRP, DAS28-ESR, CDAI, SDAI, RAPID 3) at the time of assessment. All results and outcome measures are graphed to allow visual inspection by patient and clinician. The rheumatologist can discuss these results with the patient to make shared decisions on future disease management at the time of consultation (1).

Patient-reported outcomes

A Patient Questionnaire function on the EMR allows for the collection of ePROs. This complements clinical observations from consultations. The logistics behind the ePROs are simple. The rheumatologist after discussing a questionnaire relevant to the patient's clinical situation with the patient, determines the frequency and timing of the PRO to be sent through the EMR, similar to ordering a pathology test. The questionnaire, linked to the patient's next appointment or at other defined intervals, is sent via a link through a secure server to the patient, and on completion is returned through the server to the patient's EMR, available to be accessed at the next consultation. An alternative is the PRO being answered by patients on an iPad tablet in the office waiting room and forwarded via a QR code reader to their EMR. Validated quality of life questionnaires are used to assess fatigue (FACIT-fatigue), general health and mood (Short form 1 (SF-1)) and Patient Health Questionnaire (PHQ-2)), sleep disturbance (ISI) and obstructive sleep apnoea (MAPI), Health Care Resource Utilisation, HAQ and RAPID3. Additional questionnaires have been added by OPAL to collect information in specific patients on the prevalence of skin cancers, biosimilar dispensing and demographics.

Challenges

The OPAL project has required considerable input from clinician rheumatologists in order to advise software programmers on facilitating data entry at the time of consultation without changing the flow of the doctor-patient interaction. This has been achieved particularly in RA, PsA and AS. The financial and time commitments required to establish such nuanced programming are

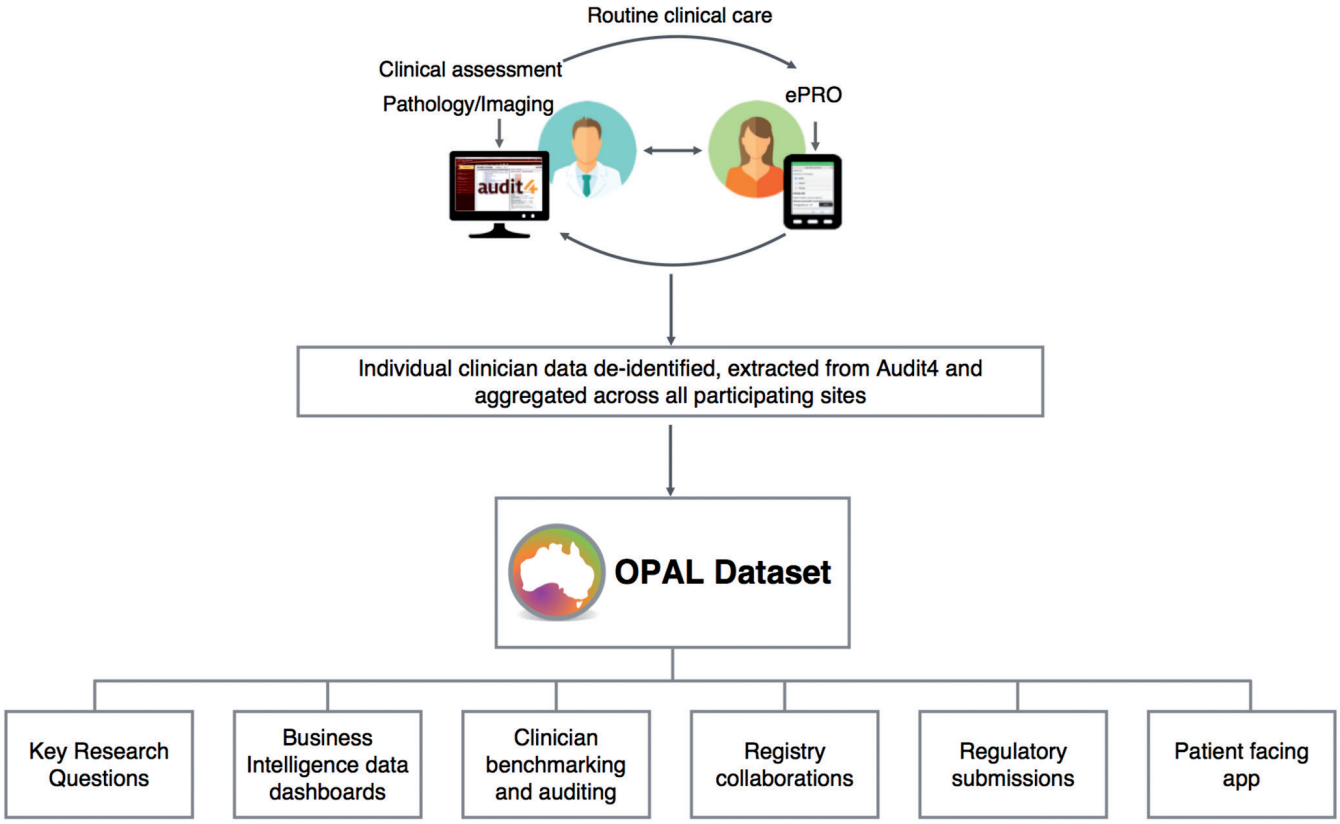


Fig. 1. Methodology used by OPAL. Data relevant to the clinical condition being treated is collected at the time of consultation and via ePROs. After de-identification and aggregation, the resulting OPAL dataset can be used for a variety of clinical care and research projects.

large and remain a constant issue as the project expands and matures. Over two million rows of code have been written for the rheumatology customisation. More workbook apps are required to optimise data collection and research in a number of other rheumatological conditions.

Because all data are collected at point of care in a real-world setting there may be gaps in data entry if the clinician is short of time or considers that a particular data entry is not relevant to the consultation. This might occur despite the EMR workbook apps having been designed to allow for logical and easy data entry. This has been addressed through group meetings and instructional educative videos to enhance use of data entry. Rates of composite outcomes, such as the DAS, have increased over time in the total OPAL group and also over time in a new OPAL member as familiarity with the programme grows.

On the other hand, the EMR has evolved to accommodate changing paradigms in rheumatological care. Treat-to-target strategies can be easily facilitated with this approach and assessment of progress of outcomes, such as remission rates, can be compared over time (2, 3). Specific questionnaires can be inserted in the workbooks to allow answers to individual issues, such as identification of barriers to establishing treat-to-target goals in RA (4).

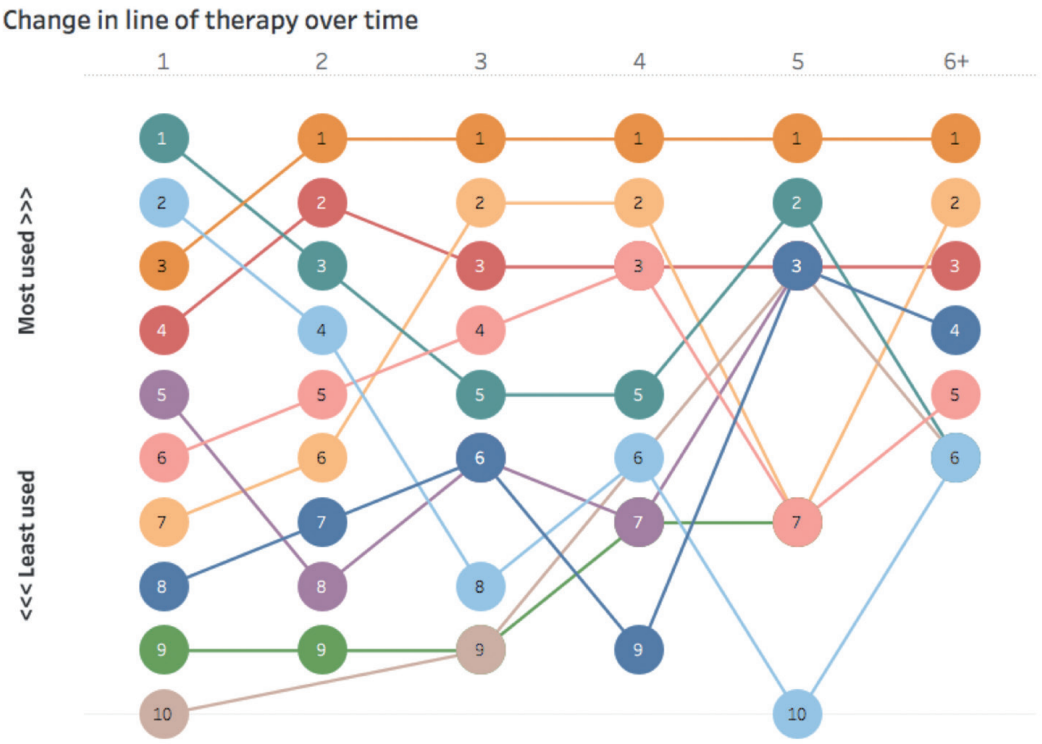
Drug interactions and adverse events can be monitored and reported (5). Drug persistence can be assessed and compared according to line of therapy and mode of action, among other variables (6).

The use of ePROs can provide more detailed information of the patients' condition in between visits to the clinician, improving the management of their condition (7). At December 2018, 16,093 suites of PROs have been sent to 3,692 patients, of which 11,456 suites were completed and returned to the EMR by 3,049 patients. The completion rate for all questionnaires sent to patients is 71%; however, 83% of patients will complete the questionnaires at least once, which indicates that the technology used to deliver the questionnaires is user friendly and patients are engaged in the management of their condition. Preliminary analyses of the quality of life data have identified a discordance between clinician assessed disease activity and patient-reported fatigue, mood and well-being (8).

Discussion

OPAL is a member-owned, not-for-profit organisation, and the data collected in the OPAL dataset are owned by the rheumatologists in the organisation. As such, there are no external conflicts of interests which may skew the collected information. Clinicians are not influenced by any financial incentives to submit data, further reducing the chance of bias and retrospective entry of data to meet the payment eligibility criteria. The Audit4 software functions as the EMR, so clinical observations are directly entered into the software as part of the regular patient consultation, and this data-gathering therefore does not impinge on routine clinical care. Data entry directly into the EMR also

A



B

bDMARD	Efficacy as monotherapy	Efficacy compared with alternatives	Mode of action/class of drug	Mode of administration	No other alternative	Safety in this patient/ comorbidities	Safety profile of drug
1	4.02%	9.90%	14.50%	6.43%	9.68%	41.28%	31.48%
2	28.61%	32.17%	25.52%	11.58%	19.35%	2.91%	4.63%
3	0.71%	1.01%	0.38%	1.93%			
4	4.26%	9.07%	10.07%	3.86%	16.13%	13.37%	14.81%
5	22.46%	24.38%	19.19%	9.00%	9.68%	16.86%	30.56%
6	1.89%	6.23%	15.07%	23.15%	6.45%	2.33%	0.93%
7		2.38%	1.58%	3.86%			
8	0.47%	2.66%	3.48%	0.32%	9.68%	16.86%	3.70%
9	24.82%	11.37%	15.14%	4.82%	6.45%	6.40%	7.41%
10	19.62%	20.35%	22.93%	36.98%	25.81%	4.07%	7.41%

Fig. 2. Business intelligence data dashboards can visually summarise OPAL big data to allow insights into numerous clinical scenarios.
For instance, in **A**, the line of therapy of each bDMARD drug in 17,000 RA patients (identified only by colour in this chart) can be plotted over successive time points to identify prescribing trends.
In **B**, the reasons for starting a particular bDMARD (represented as 1-10), can also be summarised in a dynamic and evolving fashion in each 3-month data cut.

eliminates recollection bias. The use of structured fields designated for specific data input means that information can be audited automatically, instead of searching through unstructured fields of notes. The software integrates information from each patient to a single interface, providing a longitudinal and holistic view of the patient's progress. The software provider Audit4 also facilitates the collation of de-identified patient data from each clinic into a

centralised dataset that allows for easy auditing in a uniform structure that can be uploaded in data visualisation platforms. Furthermore, the opt out consent model and completely observational nature of OPAL reduces both the burden and the risk to patients and clinicians resulting in a very low drop-out rate which is an important consideration for the longitudinal monitoring of chronic diseases.

The OPAL dataset is completely electronic, with the information stored in a secure server making it straightforward to search for specific information. Adverse events described by patients can be reported by the physician at the time of consultation, eliminating the need for reporting after data extraction. However, if a safety signal emerges in a study after de-identification, according to study-specific pre-designated Ethical approval, the OPAL safety officer can securely facilitate the identification of the patient and inform their clinician.

Considerations in the collection of big data

There are limitations in the collection of data through the point-of-care methodology that is used by OPAL. Not all relevant data may be entered into the clinical record because of time constraints or bias by the clinician. Some will enter all data and others may select data deemed to be the most relevant in the current clinical context. For instance, data always captured includes pathology results, and that related to drug prescription, but interval comorbidities, for instance, may not always be captured. The OPAL data collection is improved through targeted educational videos to help the clinician understand how to better use the EMR, as well as enhancing the patient e PROs to capture demographic and self-assessment data.

EULAR has published guidelines for collection, analysis and use of big data in rheumatic and musculoskeletal diseases (9). Overarching principles address ethical, and general principles for dealing with big data, with points to consider covering aspects of data sources, data collection, privacy by design, data platforms, data sharing and data analysis, with particular reference to artificial intelligence and machine learning. Additionally, as big data is rapidly expanding there is a need for adequate reporting of methods, benchmarking and careful data interpretation and clinical practice implementation. These approaches are evident in current big data projects (10). The OPAL programme aims to be congruent with these principles.

The role of OPAL in the rheumatology community

The OPAL dataset collates clinically relevant information from the point-of-care and is a real-world representation of the day-to-day clinical management and treatment patterns in patients with various rheumatic diseases. This information can be used in two important ways. Firstly, it can guide the development of new therapies, and minimise the risk of failed clinical trials by identifying clinically meaningful endpoints. Secondly, audit and interrogation of the clinical data in the OPAL dataset can promote quality use of medicines, thus improving patient outcomes and ensuring more efficient use of healthcare spending.

The real-world data collected through OPAL provides an insight into the effectiveness and safety of therapy as it is applied, as well as the impact of rheumatic disease on the everyday lives of the patients. This fills an unmet need in clinical research, where clinical trials are conducted in a controlled environment, with strict inclusion/exclusion criteria and monitoring, and as such may not be an accurate representation of what occurs outside of that ideal environment (11). The OPAL dataset is able to capture this information, providing a more accurate representation of the effectiveness and safety of therapies in patients with co-morbidities, as well as acting as a surrogate to reflect adherence and clinician drug preference. For example, despite the lack of comparative clinical trials we have shown that in the Australian rheumatology health-care environment the persistence of individual TNF inhibitors in inflammatory rheumatic diseases is essentially identical (6). We have also identified treatment gaps in achieving low disease activity in both RA and PsA managed in the context of the Australian health care system (2. 12). This type of data can also be used to identify patient groups who would most likely benefit from a therapy, and those who would not, increasing the chance of successful treatment for the patient, and successful clinical trials.

There were 3.9 million people in Australia affected by arthritis and other rheumatic diseases in 2015, costing the Australian health system AU$5.5 billion (13). With an aging population and increased life expectancy, the number of people affected by arthritis, along with the spending associated with these diseases, is expected to increase. Drugs targeting rheumatic diseases were the third largest therapy area worldwide, with US$53.3 billion in sales in 2016, with the prediction that this will increase to US$55.4 billion in 2022.

There is an increasing push from regulatory agencies to complement clinical trial data with real-world data. In an address to the National Academy of Sciences in September 2017, (then current) FDA commissioner Scott Gottlieb emphasised the importance of real-world data but lamented the lack of a system where "providers have the right incentives to enter clinically relevant information into EMRs at the point of care… clinically relevant information that can tell us what's happening to patients often remains in unstructured notes. We're unable to learn as much as we could about a product's profile when that information isn't accessible. We also need to find a better way to collect information directly from patients, because an EMR and claims data are really the patient's perspective filtered via the provider" (14).

A testament to the forward thinking founders of OPAL, the OPAL model of data collection was designed to address these needs, nearly 10 years and 800,000 clinical consultations before these remarks were made.

Future directions

Following patent expiration of many established biologic (b)DMARDs, many biosimilar products have been developed. There is a need to collect data

on the impact of biosimilars on patient outcomes in a real-world setting where pharmacy level brand switching may result in repeated transitions between brands of bDMARDs. A study has been initiated to investigate physician- and patient-reported disease measures following the use of anti-tumour necrosis factor-α originator (Enbrel®) and a biosimilar product (Brenzys®) in patients with RA, PsA or AS. PROs are used to confirm the receipt of originator or biosimilar products from the dispensing pharmacist, and drug persistence, efficacy and safety of switching will be compared through the analysis of patient clinical outcomes collected in Audit4.

The expansion of electronic prescribing and the ability to connect infrastructure has opened up opportunities to return dispensing information to the EMR to improve clinician awareness of medications the patient may be prescribed by other practitioners in order to monitor drug-drug interactions. This data may also be used to assess medication adherence, in particular to concomitant DMARDs and the influence on real-world clinical outcomes. Furthermore, this creates an opportunity to develop an ePRO such as a patient smartphone app to send automatic reminders to fill prescriptions and administer the medication if a dispensing notification has not been returned in the optimal time period and ascertain reasons for less than optimal adherence.

Currently, the Audit4 software used by OPAL has been specifically developed for rheumatology, but this system may be particularly useful in other chronic illnesses, which can benefit from easy access to longitudinal analyses. In Australia, chronic diseases are the leading cause of illness and disability, and the disease burden of these illnesses is expected to increase with an ageing population. Many chronic illnesses often occur together, and The Australian Institute of Health and Welfare noted that 50% of patients with RA also reported having cardiovascular disease (15). They also identified a need for "additional data on comorbidity and treatment – including on primary care; health service use; medications and whether these are being taken correctly; quality of life; and people's ability to carry out their daily lives" to get a better view of the impact of chronic diseases on people in Australia and the effectiveness of current treatment strategies. This need can be filled using the QUMI strategy applied by OPAL, with the use of custom-made EMR software, creating an integrated dataset of clinical observations for multiple diseases, an expansion of what is currently used by OPAL.

In summary, the OPAL project has shown that big data from everyday clinical rheumatological encounters can be translated into improved clinical care, audit and research in the rheumatic diseases.

Acknowledgements

OPAL Rheumatology gratefully acknowledges Roche Australia for provision of an unrestricted enabling grant to initiate this project.

References

1. TYMMS K, LITTLEJOHN G. OPAL: a clinician driven point of care observational data management consortium. *Clin Exp Rheumatol* 2014; 32 (Suppl. 85): S150-52.
2. LITTLEJOHN G, ROBERTS L, ARNOLD M *et al*.: A multi-center, observational study shows high proportion of Australian rheumatoid arthritis patients have inadequate disease control. *Int J Rheum Dis* 2013; 16: 532-8.
3. LITTLEJOHN G, ROBERTS L, BIRD P *et al*.: Patients with rheumatoid arthritis in the Australian OPAL cohort show significant improvement in disease activity over 5 years: a multicenter observational study. *J Rheumatol* 2015; 42: 1603-9.
4. TYMMS K, ZOCHLING J, SCOTT J *et al*.: Barriers to optimal disease control for rheumatoid arthritis patients with moderate and high disease activity. *Arthritis Care Res* (Hoboken) 2014; 66: 190-6.
5. BIRD P, GRIFFITHS H, TYMMS K *et al*.: The SMILE study -- safety of methotrexate in combination with leflunomide in rheumatoid arthritis. *J Rheumatol* 2013; 40: 228-35.
6. TYMMS K, LITTLEJOHN G, GRIFFITHS H *et al*.: Treatment patterns among patients with rheumatic disease (rheumatoid arthritis (RA), ankylosing spondylitis (AS), psoriatic arthritis (PsA) and undifferentiated arthritis (UnA)) treated with subcutaneous TNF inhibitors. *Clin Rheumatol* 2018; 37: 1617-23.
7. ROTENSTEIN LS, HUCKMAN RS, WAGLE NW: making patients and doctors happier - the potential of patient-reported outcomes. *N Engl J Med* 2017; 377: 1309-12.
8. BIRD P, LITTLEJOHN G, GRIFFITHS H *et al*.: Fatigue, poor health and mood disturbance persist in rheumatoid arthritis and psoriatic arthritis patients despite disease remission: The OPAL DEEPER study. *Arthritis Rheumatol* 2017; 69 (suppl 10):1354.
9. GOSSEC L, KEDRA J, SERVY H *et al*.: EULAR points to consider for the use of big data in rheumatic and musculoskeletal diseases. *Ann Rheum Dis* 2020; 79: 69-76.
10. ACAR-DENIZLI N, KOSTOV B, RAMOS-CASALS M, ON BEHALF OF THE SJOGREN BIG DATA CONSORTIUM: The Big Data Sjögren Consortium: a project for a new data science era. *Clin Exp Rheumatol* 2019; 37 Suppl. 118: 19-23.
11. KREMER JM: The Corrona US registry of rheumatic and autoimmune diseases. *Clin Exp Rheumatol* 2016; 34: S96-S9.
12. TYMMS K, KELLY A, BIRD P *et al*.: Psoriatic arthritis treatment regimens, therapy duration and reasons for cessation in the biologics era: a multi-centre Australian study. *Int J Rheum Dis* 2018; 21: 510-6.
13. Counting the cost, part 1: healthcare costs, Arthritis Australia, Sydney.
14. GOTTLIEB S: 2017. Remarks by Dr. Gottlieb to the National Academy of Sciences on the impact of real world evidence on medical product development. . https://www.fda.gov/newsevents/speeches/ucm576519.htm
15. Rheumatoid Arthritis. *Rep Web 62*, 2017, Australian Institute of Health and Welfare.