

# You and your EMR: the research perspective

## Part 2. How structure matters

Bridget L. Ryan MSc PhD Joshua D. Shadd MD CCFP Amanda Terry MA PhD Sonny Cejic MSc MD CCFP  
 Vijaya Chevendra MSc Amardeep Thind MD PhD

*When we last saw Dr Park,<sup>1</sup> she and her team had chosen their electronic medical record (EMR) and were implementing their new system. They had developed prototype questions in order to test-drive their EMR. Throughout the implementation process, they became increasingly aware that if they wanted to answer their questions, they needed to accurately and reliably input and retrieve relevant information about their patients.*

Inputting information into your EMR requires a balance between being efficient and being complete and accurate. Both goals are critical for patient care. When you intend to conduct research using data from your EMR, there is an additional goal of being able to retrieve the data in a reliable and consistent manner. Use your prototype questions as a guide to decide *what* information is critical for your research, *where* you will put this information, and *how* you will structure it.


Let's look at a prototype example. You decide that *what* you want to be able to identify is your patients who smoke. You decide that *where* you will record smoking is in the cumulative patient profile. In order to retrieve this information, you decide *how* everyone in

the practice will enter this information in a consistent manner. Everyone will indicate "smoker" in a structured field in the cumulative patient profile rather than entering, for example, "smokes" or "currently smoking" in an unstructured way. Therefore, to consistently retrieve all patients who smoke, everyone must use the cumulative patient profile and pick the same term for "smoker." Consistency, then, relates both to where and how you enter information.

### Structured and unstructured data

It is worth looking in more detail at how information is entered. This is primarily a choice between unstructured and structured data. **Figure 1** outlines the types of data along this continuum. Unstructured data include scanned documents from which, with current technology, it is virtually impossible to efficiently extract data. Clinical notes are usually unstructured, as most consist of free text or narrative. Narrative in family practice is important to patient care and can be appealing because it emulates a paper chart and does not require navigating from one field to another within the EMR. However, it is difficult to retrieve aggregated patient information from narrative

**Figure 1. Taxonomy of EMR data structure**

Level of structure	Type of data	Examples
Unstructured data  	Scanned documents	Historical paper records Laboratory results Consultation letters Images
	Narrative fields	Clinical notes
	Free-format fields No conventions	Multiple ways of entering condition (eg, "diabetes," "DM," or "diabetic")
	Free-format fields Conventions	Consistent entry for condition (eg, "diabetes" not "DM" or "diabetic")
	Structured fields	Drop-down lists Radio on and off buttons Date-formatted fields
Structured data		

DM—diabetes mellitus, EMR—electronic medical record.

data.<sup>2</sup> Therefore, it is ideal to minimize the narrative and use structured fields wherever possible.


Structured data can take various forms. You will have available, and will likely use, different methods for different types of information. Less-structured fields are free-text fields with little or no built-in structure. The challenge here is for everyone to enter information in a consistent manner. You should ensure that people agree to data conventions. If not, you will need to know every way (including spelling mistakes) that people have entered data (eg, to indicate diabetes, "type 2 diabetes," "DM," [diabetes mellitus] or "diabetic" are all options). Highly structured fields include radio on and off buttons, checklists, and text fields with defined drop-down lists. Here the data will be consistent as long as everyone makes the same choices to represent the same information in the same place. The need to check and click numerous boxes, however, might impede efficiency. Furthermore, you might not always find exactly what you want among the pre-existing choices. Paying attention to the number of key strokes or clicks required to enter data will help inform how much you can expect from people entering information.

Coding schemes and templates can be employed to aid in structuring data. Coding schemes such as the codes for International Classification of Diseases, ICD-9, ICD-10, or ICPC, can be used to code reasons for encounters, diagnoses, and problems.<sup>3</sup> Templates that mimic paper flow sheets can be used to organize data around common topic areas such as vaccination records or diabetes.

### Keys to consistency

The keys to consistency are training and monitoring over time. Ensure that everyone agrees to what information will be input and where and how it will be entered. You should periodically validate your data by conducting queries to retrieve information to determine if what you retrieve is what you expect. You should provide feedback to all your EMR users to collectively review whether or not you are getting consistent information. These tests can point to further processes that you need to establish or further training that might be beneficial. The information you wish to retrieve might change over time—the EMR can evolve; just be aware that this

might affect the quality of the data and the time frame in which it can be retrieved.

As EMR technology evolves, structuring of data will become more transparent and more easily accomplished. Future technology might offer flexibility so that much of the structuring of data is happening behind the scenes. In the meantime, learning to use your EMR in a more structured way can involve a considerable learning curve, and data entry can be initially less efficient; however, mastering accurate and consistent entering of data opens up many possibilities for improving both patient care and research. 

**Dr Ryan** is a postdoctoral fellow and **Dr Shadd** is Assistant Professor, both in the Centre for Studies in Family Medicine in the Department of Family Medicine at the Schulich School of Medicine and Dentistry at the University of Western Ontario (UWO) in London. **Dr Terry** is Assistant Professor in the Centre for Studies in Family Medicine in the Department of Family Medicine and the Department of Epidemiology and Biostatistics at the Schulich School of Medicine and Dentistry at the UWO. **Dr Cejic** is a family physician in an academic practice in London and is Associate Clinical Professor in the Department of Family Medicine at the UWO. **Ms Chevendra** is IT Consultant in the Centre for Studies in Family Medicine at the UWO. **Dr Thind** is Associate Professor in the Centre for Studies in Family Medicine in the Department of Family Medicine and the Department of Epidemiology and Biostatistics at the Schulich School of Medicine and Dentistry at the UWO.

### Competing interests

**Dr Cejic** is Chair of the Nightingale Advisory Board at Nightingale Informatix (the vendor of the electronic medical record that **Dr Cejic** uses). None of the other authors has any competing interests to declare.

### Correspondence

**Dr Bridget L. Ryan**, Centre for Studies in Family Medicine, Schulich School of Medicine and Dentistry, Department of Family Medicine, University of Western Ontario, Suite 245, 100 Collip Circle, London, ON N6G 4X8; telephone 519 661-2111, extension 20115; fax 519 858-5029; e-mail [bryan@uwo.ca](mailto:bryan@uwo.ca)

### References

1. Ryan BL, Cejic S, Shadd JD, Terry A, Chevendra V, Thind A. You and your EMR: the research perspective. Part 1. Selecting and implementing an EMR. *Can Fam Physician* 2011;57:1090-1.
2. Terry AL, Chevendra V, Thind A, Stewart M, Marshall JN, Cejic S. Using your electronic medical record for research: a primer for avoiding pitfalls. *Fam Pract* 2010;27(1):121-6. Epub 2009 Oct 14.
3. World Health Organization [website]. *International Classification of Diseases (ICD)*. Geneva, Switz: World Health Organization; 2011. Available from: [www.who.int/classifications/icd/en/](http://www.who.int/classifications/icd/en/). Accessed 2011 Oct 20.

---

Hypothesis is a quarterly series in *Canadian Family Physician*, coordinated by the Section of Researchers of the College of Family Physicians of Canada. The goal is to explore clinically relevant research concepts for all CFP readers. Submissions are invited from researchers and nonresearchers. Ideas or submissions can be submitted online at <http://mc.manuscriptcentral.com/cfp> or through the CFP website [www.cfp.ca](http://www.cfp.ca) under "Authors and Reviewers."

\*\*\*