

A Framework and Standardized Methodology for Developing Minimum Clinical Datasets

Piper A. Svensson-Ranallo, MHI¹, Terrence J. Adam, MD, PhD^{1,2}, François Sainfort, PhD^{1,3}

¹Institute for Health Informatics, University of Minnesota, Minneapolis, MN; ²College of Pharmacy, University of Minnesota, Minneapolis, MN; ³School of Public Health, University of Minnesota, Minneapolis, MN

Abstract

The concept of the minimum dataset (MDS) is taking on an increasingly important role in healthcare. In the current environment of health information exchange and universal implementation of electronic health records, work related to the development of one specific type of MDS, the minimum clinical dataset (MCDS), is beginning to permeate the literature. While there is currently no unified definition of either an MDS or an MCDS, an MDS is generally agreed to be a coherent set of explicitly defined data elements. Despite the growing body of literature on MCDSs, very little empirical evidence exists in the literature related to best methods for developing them. The primary objective of the current study is to fill this gap. By presenting a streamlined approach to the development of MCDSs the current study attempts to provide individuals and organizations with a coherent methodology and framework for developing a high quality MCDS.

Introduction

The term ‘*minimum dataset*’, or MDS, is a commonly used, but poorly defined term in the healthcare literature. Conceptualizations of the MDS range from that of an essential¹ or pertinent set of data elements

related to a single clinical condition,^{2,3} procedure,^{4,5} specialty,⁶⁻⁸ discipline^{1,9} or healthcare process,^{10,11} to that of a comprehensive and inclusive set of data elements related to an entire domain of healthcare (e.g.: the United States’ Long Term Care *Minimum Dataset*,¹² the UK’s *Mental Health Minimum Dataset*,¹³ and New Zealand’s *General Practice Minimum Dataset*¹⁴).

This paper focuses on the subset of MDSs developed for collecting data during the routine process of care: *the minimum clinical dataset*, or MCDS. Using Berwick’s¹⁵ framework of quality, we define a MCDS as an MDS developed for, used by, and targeting actions that occur at the ‘microsystem’ level of healthcare. According to Berwick, quality can be achieved by addressing processes at four levels: that of the patient (level A); that of the microsystem, or small units of care delivery (level B); that of the organization (level C); and that of the larger physical, social, economic and political environment (level D).

For the current study, we define an MCDS as a) a coherent, explicitly articulated set of standardized data elements; b) developed using an explicit, empirically based approach to defining and naming relevant clinical constructs; c) designed to optimally

Table 1: Differences between a *Minimum Dataset* and a *Minimum Clinical Dataset*

	Minimum Dataset (MDS)	Minimum Clinical Dataset (MCDS)
Primary Objectives	Provision of the highest quality of care as defined by <i>population</i> averages, and constrained by the need to balance multiple-stakeholder objectives	Provision of personalized, high quality care as defined by the ability to achieve the outcomes desired by individual patients
Construct Focus	Constructs related to <i>organizations</i> and/or <i>systems</i> (primarily Berwick’s level C,D)	Constructs related to <i>patient</i> and healthcare <i>microsystem</i> (primarily Berwick’s level A, B)
Data Collection	Data is rarely collected solely as part routine delivery of care; typically the data collection process is MDS-specific	Data is collected, used, and analyzed at the <i>microsystem</i> level for routine care processes
Data Source	Multiple sources from all levels of healthcare system (clinical, operational, organization)	<i>Patient</i> and <i>microsystems</i> that interface directly with patient
Data Use	<i>Healthcare organization</i> and <i>environment</i> (primarily Berwick’s level C,D)	<i>Patient/community</i> and <i>microsystem</i> (primarily Berwick’s level A, B)

represent and capture data at the patient-*microsystem* interface; d) implemented in such a way that it can be integrated with related MDSs and/or MCDSs; and e) oriented towards the acquisition of actionable knowledge to be used at the microsystem level.

Our concept of the MCDS is very similar to what Wirtschafter & Mesel¹⁶ described more than 30 years ago as a “minimum care assurance data set”. In their paper, the authors defined this dataset as those data elements that focus clinical attention on the variables most relevant to achieving predefined clinical objectives, facilitating ongoing analysis of outcomes, and supporting timely corrective action in response to specific clinical results.

Methods

Data Collection

We used a key-word search methodology to identify the body of literature in which some variation on the term ‘minimum dataset’ was used. To do this, we searched PubMed for publications in which the terms ‘*minimum*’, ‘*data*’ and ‘*set*’ or the terms ‘*minimum*’ and ‘*dataset*’ occurred together in either the title or the abstract. All publications indexed between January 1, 1950 and January 1, 2011 were included in the study.

We began by assessing the different concepts to which authors applied the term ‘minimum dataset’. We then classified all publications into the two high-level categories of MDS or MCDS. To develop an explicit, standardized methodology for creating an MCDS, we analyzed the subset of MCDS publications in which development of the MCDS was described. We identified the core structural and functional attributes of the MCDS, and critically evaluated the development methods used by the authors. In addition, we analyzed information from the results, discussion and conclusions of these publications regarding authors’ perceptions, with the goal of identifying the methods most likely to produce high quality MCDSs.

Results

A total of 3208 articles were identified. Of these, 177 (5.5%) had the term ‘*minimum dataset*’ in the title. A total of 1601 (49.9%) articles were excluded either because the term ‘*minimum*’ was not used to qualify the term ‘*dataset*’ in the abstract or title; it was a duplicate entry; or the minimum dataset described was not directly related to healthcare (e.g.: a *minimum dataset* of soil quality variables). Of the remaining 1607 publications, 366 articles or (22.8%) were classified as describing *minimal clinical datasets* as defined by the current study. The

remaining 1241 articles (77.2%) were classified as describing general *minimum datasets*.

There was a great deal of variability within the general MDS category. The term *national minimum dataset*, for example, was frequently used to describe a number of distinct types of datasets and registries. These datasets were developed and used for purposes ranging from surveillance^{17,18} and epidemiological tracking¹⁹ to service planning,²⁰ budgeting,²¹ and population-level clinical research.²²

Not only were many different types of datasets described as MDSs, a wide range of organizations and entities have developed healthcare related MDSs. These include international organizations; multinational coalitions; national and local governments; professional organizations and entities; and both clinical and research organizations.

In addition to the types of minimum datasets described, publications were also assessed for the clinical specialty area upon which they focused. As shown in figure 1, the three specialty areas for which the largest number publications exist are geriatrics (26%), oncology (23%), and pathology (20%).

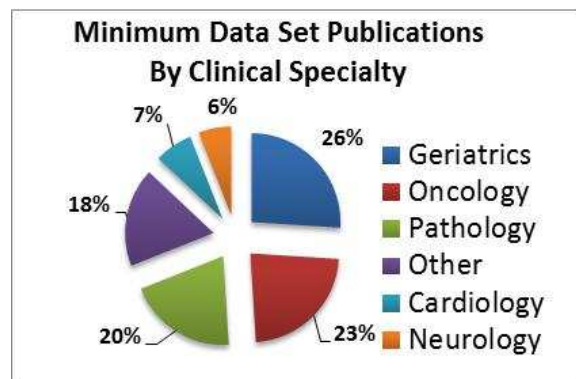


Figure 1. Publications by *Minimum Dataset* Clinical Specialty.

Methodology for Developing Minimum Clinical Datasets

The methods commonly used by MCDS developers ranged from the use of hired consultants,⁸ formation of expert⁹ or representative⁸ stakeholder committees, stakeholder interviews,²³ and distribution of surveys;¹¹ to systematic literature reviews, chart reviews, and reviews of both existing clinical information systems⁹ and clinical data collection tools.²⁵ In a handful of studies, MCDS developers employed formal methodologies, such as the Delphi technique,¹⁰ for achieving consensus among content experts.

Consistent with approaches reported by many MCDS developers, we propose a bottom-up, multi-modal approach in which data elements identified in both the literature and patient charts are critically evaluated by domain experts through a formal and iterative process (figure 2). The chart review process is a core component of this approach, as it is expected to provide insight into many *implicit representations* of clinical constructs that may not exist in the literature. The systematic review of the literature ensures that the set of data elements considered for inclusion is the comprehensive, and not simply the currently or commonly used, set of elements. The systematic review also allows for the weighting of evidence regarding the potential relevance of a given data element to the core dataset.

When this multi-modal method is used to identify the pertinent data elements for the dataset, results generated using each method must be harmonized (i.e.: translated into a common language). The process of harmonizing and naming disparate representations of concepts - both within and between data source types (i.e.: the literature, patient records, and expert opinion) - is expected to facilitate the identification of new or potentially confounded constructs.

Once the relevant clinical constructs and related variables have been identified and operationally defined, the set of possible terms used for naming the data elements must also be finalized. Standard nomenclatures (such as SNOMED-CT) should be used to select both *preferred* and *fully specified terms* for each concept. Any discrepancy between the data elements representing clinical entities and attributes on the one hand, and their representation (or existence) in standard classification systems or nomenclatures on the other should be reconciled (i.e.: by submission of the concept or code to the appropriate concept representation system). This process allows for the creation of elements and attributes that can be incorporated into standard representation systems for future use in structured instrument development.

Discussion

While the term MDS is commonly thought to describe an essential, uniform set of data elements to be collected across time and organizations,¹ the current analysis suggests that the term is also widely used in healthcare to describe an ontology;²⁴ an existing set of data elements used for a specific purpose;²⁵ and a standardized protocol for collecting data.²⁶ In addition, the term is sometimes used to describe an entirely different set of constructs, including the minimum number of data points

required for an adequate logistic regression model; the specific, minimum number and locations of anatomically placed electrodes required in specific imaging techniques; or the number of evidence based practices used during treatment of a set of patients.¹³

While the current analysis highlights the multitude of constructs to which the term ‘minimum dataset’ has been applied, it is also clear that one specific, and clinically salient type of dataset can be identified, and clearly distinguished from other types of MDSs found in the literature: the *minimum clinical dataset*.

The MCDS has important implications not only for decision making in clinical care, but also for workflow management and reimbursement policies that ultimately support clinicians in performing the activities known to be associated with quality outcomes. For example, clinicians routinely perform a systematic, highly granular set of (often cognitive) activities during the routine provision of care. Incorporating these activities and elements into a MCDS provides clinicians not only with subtle decision support (i.e.: an order-set for essential data to consider and collect), but also with an efficient mechanism for capturing the results of these findings. Furthermore, the MCDS data collection process itself represents a clinical order set (for data collection) and produces a tangible clinically-relevant product (information) subject to valuation and pricing.

There are several limitations to our findings. One limitation of the keyword search method was that it was based on a relatively narrow set of terms. A more extensive keyword search would include the terms ‘minimal’, ‘uniform’, and ‘core’. Similarly, a limitation of the proposed methodology is that it was developed based on publications identified through a preliminary, rather than systematic, review of the literature. It’s possible that publications describing a specific approach to the development of *minimum clinical datasets* were not captured using the keyword search methodology.

Conclusions

The current study reviewed the literature related to *minimum datasets* (MDS), with a particular focus on a rapidly growing type of MDS: the *minimum clinical dataset* (MCDS). To address the existing knowledge gap regarding optimal approaches for developing MCDSs, we reviewed, assessed, and harmonized a number of methodological approaches described in the literature. We developed and proposed a streamlined methodology for developing *minimum clinical datasets*. Using the framework that served as the basis of the IOM’s landmark “Quality Chasm” report, and made explicit by Berwick in his “User’s

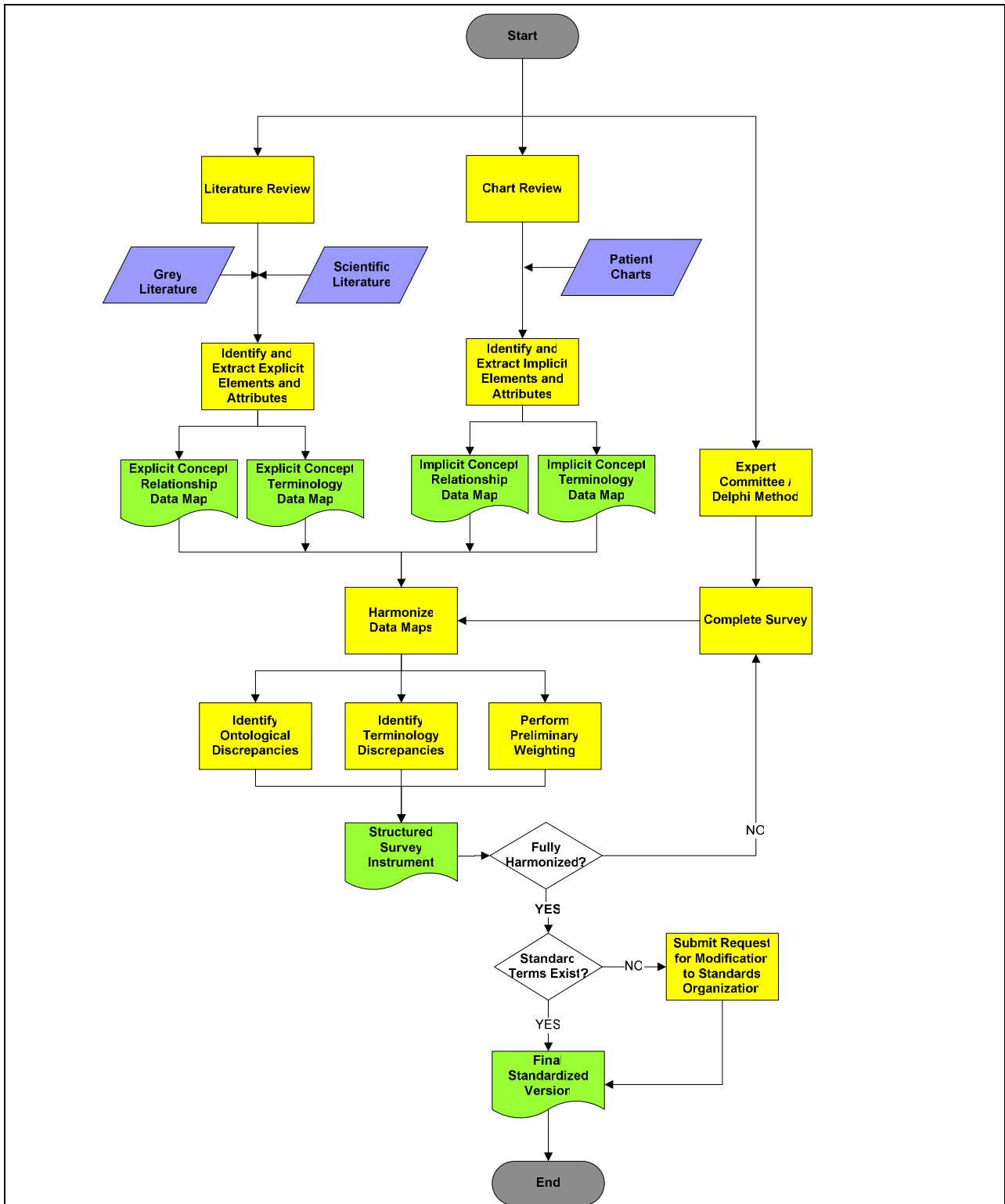


Figure 2: Development Methodology for the First Iteration of a *Minimum Clinical Dataset*

Manual for the IOM Report”, we define the *minimum clinical dataset* as a critical component of the healthcare delivery system. As a clinical tool, a well-designed MCDS is essential not only to ensure the

delivery of high quality care to individual patients, but also to facilitate the collection of high-value clinical data necessary for the acquisition of new and better clinical knowledge. The primary function of

the MCDS is to support optimal decision making at the point of care - what Berwick refers to as the "true north" of healthcare quality - where patient experience intersects with microsystems of care to produce quality outcomes.

References

1. Werley HH, Devine EC, Zorn CR. Nursing needs its own minimum data set. *Am J Nurs*. 1988;88(12):1651-1653.
2. Leong SC, White PS. Outcomes following surgical decompression for dysthyroid orbitopathy (graves' disease). *Current Opinion in Otolaryngology & Head and Neck Surgery* 2010; 18(1); 37-43.
3. McCormick J, Sims EJ, Green MW, Mehta G, Culross F, Mehta A. Comparative analysis of Cystic Fibrosis Registry data from the UK with USA, France and Australasia. *J Cyst Fibros*. 2005;4(2):115-122.
4. Chan W, Clark DJ, Ajani AE, et al. Progress Towards a National Cardiac Procedure Database-Development of the Australasian Society of Cardiac and Thoracic Surgeons (ASCTS) and Melbourne Interventional Group (MIG) Registries. *Heart Lung Circ*. 2010.
5. Nixon I, Kunanandam T, Mackenzie K. Minimum dataset for endolaryngeal surgery: pilot study. *J Laryngol Otol*. 2010;124(9):980-985.
6. Thomas K, Emberton M, Mundy AR. Towards a minimum dataset in urology. *BJU Int*. 2000;86(7):765-772.
7. Lack JA, Stuart-Taylor M, Tecklenburg A. An anaesthetic minimum dataset and report format. Society for Computing and Technology in Anaesthesia (SCATA). European Society for Computing and Technology in Anaesthesia (ESCTAIC). *Br J Anaesth*. 1994;73(2):256-260.
8. Bean KB. Development of the Society of Gastroenterology Nurses and Associates Minimum Data Set: an evidence-based resource. *Gastroenterol Nurs*. 2005;28(1):56-58.
9. Ireland RS, Jenner AM, Williams MJ, Tickle M. A clinical minimum data set for primary dental care. *Br Dent J*. 2001;190(12):663-667.
10. Bagley TC, Schaffer, J. Minimum data set development: Air transport time-related terms. *International Journal of Medical Informatics* 2002; 65(2);121-133.
11. Mistry NK, Toulany A, Edmonds JF, Matlow A. Optimizing physician handover through the creation of a comprehensive minimum data set. *Healthc Q*. 2010;13 Spec No:102-109.
12. Morris JN, Hawes C, Fries BE, et al. Designing the national resident assessment instrument for nursing homes. *Gerontologist*. 1990;30(3):293-307.
13. Glover GR, Sinclair-Smith H. Computerised information systems in English mental health care providers in 1998. *Soc Psychiatry Psychiatr Epidemiol*. 2000;35(11):518-522.
14. Hall J, Tomlin A, Martin I, Tilyard M. A general practice minimum data set for New Zealand. *N Z Med J*. 2002; 115(1163); U200.
15. Berwick, DM. A user's manual for the IOM's "Quality Chasm" report. *Health Affairs*, 2002; 21; 80-90.
16. Wirtschafter DD, Mesel E. A strategy for redesigning the medical record for quality assurance. *Med Care*. 1976;14(1):68-76.
17. Madsen M, Gudnason V, Pajak A, et al. Population-based register of acute myocardial infarction: manual of operations. *Eur J Cardiovasc Prev Rehabil*. 2007;14 Suppl 3:S3-22.
18. Irving LM, Norton RN, Langley JD. Injury surveillance in public hospital emergency departments. *N Z Med J*. 1994;107(979):222-223.
19. Wong A, Taylor DM, Ashby K, Robinson J. Changing epidemiology of intentional antidepressant drug overdose in Victoria, Australia. *Aust N Z J Psychiatry*. 2010;44(8):759-764.
20. Middleton S, Gardner G, Gardner A, Della P, Gibb M, Millar L. The first Australian nurse practitioner census: A protocol to guide standardized collection of information about an emergent professional group. *Int J Nurs Pract*. 2010;16(5):517-524.
21. Miller P, Parkin D, Craig N, Lewis D, Gerard K. Less fog on the Tyne? Programme budgeting in Newcastle and North Tyneside. *Health Policy*. 1997;40(3):217-229.
22. Blondel B, Zein A, Ghosn N, du Mazaubrun C, Breart G. Collecting population-based perinatal data efficiently: the example of the Lebanese National Perinatal Survey. *Paediatr Perinat Epidemiol*. 2006;20(5):416-424.
23. Evans SM, Murray A, Patrick I, Fitzgerald M, Smith S, Cameron P. Clinical handover in the trauma setting: a qualitative study of paramedics and trauma team members. *Qual Saf Health Care*. 2010;19(6):e57.
24. Parkinson H, Aitken S, Baldock RA, et al. The SOFG Anatomy Entry List (SAEL): an annotation tool for functional genomics data. *Comp Funct Genomics*. 2004;5(6-7):521-527.
25. Crawley C, Iacobelli S, Bjorkstrand B, Apperley JF, Niederwieser D, Gahrton G. Reduced-intensity conditioning for myeloma: lower nonrelapse mortality but higher relapse rates compared with myeloablative conditioning. *Blood*. 2007;109(8):3588-3594.
26. James J, Evans JA, Young T, Clark M. Pressure ulcer prevalence across Welsh orthopaedic units and community hospitals: surveys based on the European Pressure Ulcer Advisory Panel minimum data set. *Int Wound J*. 2010;7(3):147-152.