



Research Article

# Strengthening the Foundation of General Practice Evidence in Ireland by Addressing the Data Quality Issues in a Structured Secondary Prevention Programme for Cardiovascular Disease

Claire Collins<sup>1</sup>, Caitriona Finn<sup>2</sup>, Brian Meade<sup>3</sup> and Fionan O'Cuinneagain<sup>4</sup>

<sup>1,2,4</sup> Irish College of General Practitioners, Dublin, Ireland

<sup>3</sup>Kilmacud Medical Centre, Dublin, Ireland

Correspondence should be addressed to: Claire Collins; [claire.collins@icgp.ie](mailto:claire.collins@icgp.ie)

Received date: 23 October 2013; Accepted date: 21 January 2014; Published date: 12 May 2014

Academic Editor: Emer R. Mcgrath

Copyright © 2014. Claire Collins, Caitriona Finn, Brian Meade and Fionan O'Cuinneagain. Distributed under Creative Commons CC-BY 3.0

## Abstract

The Heartwatch Programme provides a structure and protocol for the continuing care of patients for the secondary prevention of cardiovascular disease in general practice in Ireland. The database consists of 17,399 patients and 185,855 consultations. The Independent National Data Centre (INDC) receives data from the participating practices and is responsible for data management and report production. Some of the data quality issues identified resulted because the concept of evaluation had not been fully taken on board at the commencement of the programme, and a data quality management process was not instituted from the outset. Strategies for managing and improving data quality were developed and the system design and protocols enhanced. The INDC system features full automation of the data processing to ensure it meets the agreed data quality targets; it features online facilities for both participating practices and the central administration to upload, check and correct data in addition to running financial reports and pre-defined and customized GP, regional and national demographic and clinical reports. The programme has involved considerable change management within general practice which has had far reaching benefits in many areas in addition to coronary heart disease prevention. As noted internationally, there is substantial potential to capitalize on the economy of scale benefits to establish other healthcare programmes and projects which also necessitate reliable and valid data capture from general practice. The experience and learning about data quality from general practice, a large sector of the health care service, will facilitate other structured care programmes in this sector and throughout the healthcare environment.

**Keywords:** General practice; data quality; information and communication technology (ICT); secondary prevention of cardiovascular disease.

---

**Cite this Article as:** Claire Collins, Caitriona Finn, Brian Meade and Fionan O'Cuinneagain (2014), "Strengthening the Foundation of General Practice Evidence in Ireland by Addressing the Data Quality Issues in a Structured Secondary Prevention Programme for Cardiovascular Disease", JMED Research, Vol. 2014 (2014), Article ID 315020, DOI: 10.5171/2014.315020

## Introduction

*Data are of high quality if they are fit for their intended use in operations, decision-making, and planning (Juran, 1964).*

The Heartwatch Programme provides a structure and protocol for the continuing care of patients for the secondary prevention of cardiovascular disease in general practice/family medicine in Ireland. The programme targets 20% of general practices with patients seen on a quarterly basis and care implemented according to defined clinical protocols (The National Heartwatch Programme, 2004). Heartwatch is the largest database on cardiovascular disease in general practice in Ireland with 17,399 patients and 185,855 consultations. There has been substantial international interest in the programme with numerous requests to outline and discuss the programme approach and strategies from international colleagues at their regional conventions.

The health benefits of this programme have been documented (The National Heartwatch Programme, 2004 and 2006; McGrath et al, 2012), and action taken with regard to areas where health and lifestyle improvements were not shown to be achieved (Lambe and Collins, 2010). However, although this was the intended purpose of the programme, this analysis and reporting does not convey the data management processes involved in ensuring the data to be 'fit for use'.

Fitness for use is seen as an important aspect of data quality (Madnick et al., 2009; US Census Bureau, 2006; Chrisman, 1991). Redman (2001) suggested that for data to be fit for use they must be accessible, accurate, timely, complete, consistent with other sources, relevant, comprehensive, provide a proper level of detail, be easy to read and easy to interpret.

This paper describes the methods employed to monitor and address data quality issues in order to produce a large scale quality assured

database from Irish general practice. It outlines the approach taken to ensure effective and acceptable governance of the information system in addition to a range of solutions for data errors and omissions where data is being collected during routine patient consultations. With the increasing use of electronic recording in healthcare settings and the ease of data collection and onward transmission to central databases and registries, such rigorous attention to data quality is required, and hence the solutions outlined here are relevant to other contexts where healthcare providers are entering data on aspects of patient care.

## Data Quality Management Processes

The initial implementation phase of the programme employed a standardised approach, adhered to internationally recognised cardiovascular prevention guidelines and followed defined clinical care protocols, which included the recording of specified data.

A national programme centre (NPC), was set up to implement the programme, and an Independent National Data Centre (INDC) was established which received the data from the participating practices, and distributed aggregated anonymised relevant data reports to applicant agencies and organisations.

A national steering committee (NSC) oversaw the implementation of the Heartwatch Programme and, was made up of representatives of all of the major stakeholders. A data management committee (DMC) oversaw the activities of the INDC and reported to the NSC. The DMC was responsible for data quality assurance and monitoring, and for providing permission for data access. Demographic and clinical reports were produced on approval by this committee.

The four main general practice (GP) software suppliers in Ireland at the time of commencement of the Heartwatch Programme formed a Health Informatics

Association, and each of these providers made available a Heartwatch system module for GP users to integrate with their current practice software. Prior to the availability of this integrated software, an interim software commissioned for the programme by the INDC was utilised by practices.

A detailed non-technical specification document of all data fields with detailed instructions and explanations were provided to participating practices and training and ongoing support provided locally. File generation schema and software architecture documents were produced to inform the software providers of the data requirements.

One year after commencement, under the direction of the DMC, an external company was contracted to undertake a quality assessment of the data collection, cleaning and analysis processes conducted within the INDC.

The data quality improvement approach followed that recommended by Madnick and

Wang (1992) with the cycles of Define, Measure, Analyze, and Improve.

Data profiling is the use of analytical techniques on data for the purpose of developing a thorough knowledge of its content, structure and quality. It is a process of developing information about data instead of information from data, which involves the following steps:

- Collect documentation (non-technical specification, file generation schema)
- Review the data itself (individual patient data entered by GPs and returned to the INDC where it was merged into a central database)
- Compare data to documentation
- Identify and detail specific issues.

(DeMaio, 2002)

On completion of the data profiling, the following data quality issues (with examples) were identified:

Labelling inconsistencies	Different value labels produced by software packages e.g. MALE, male, Male
Inconsistent data types	Weight circumference was produced as a text field from one software package
Out of range values	Impossible clinical measures e.g. SBP of 310
Invalid values	Date of patient registration prior to programme commencement in 2003
Missing values	Mandatory clinical measurements not recorded at visits
Unused fields	Under-reporting of events such as hospitalisation
Rule exceptions	If a clinical measure did not require intervention, it was not re-measured at the subsequent visit. If it was measured, need to establish if the previous entry was incorrect

Default value errors	If a clinical measure was not re-measured at the subsequent visit, a specific default value was recorded. Errors included use of the incorrect default value and use of the default value when re-measurement should have taken place
Duplicates	Apparent duplication of data for patients – double entry or visit date error

The issues identified were traced back to the practice software (first two issues) and the data source at the time of data entry (all other issues). Hence, as advised by Madnick et al (2009) following the assessment and evaluation of data quality, both information system and protocol changes were adopted to address the issues identified.

Improvements were obtained through the use of:

1. Automatic correction of labelling errors prior to merging into the central database
2. Automatic correction of field types prior to merging into the central database
3. Inclusion of the data fields in a national process of accreditation of all GP software providers
4. Setting of out of range values at data source level
5. Automatic return of flagged errors/inconsistencies/missing data to the data source
6. Automatic rejection of apparent duplicates to data source
7. In-built checking of default values and immediate return of queries to data source
8. Noted limitations of the use and interpretation of data in unused fields

9. Linking of payment to fully completed and accurate data return.

The introduction of the above measures has eliminated the identified data quality issues in the final database, with the exception of the unused data fields. The reasons for lack of use could be identified, however, no workable solution was available to address this for certain instances, and users are notified of the limitations of these fields.

Two independent reviews of the Heartwatch data have been conducted with the following conclusions: “It is commendable that Heartwatch managed to become operational within a relatively short period of time, and that systems were developed quickly to facilitate the electronic interchange of Heartwatch data. Many of the problems reported with regard to software bugs and datasets during the early implementation period of Heartwatch are highly typical of new projects, and were rectified once the problems had been identified” (Capita Consulting 2005).

“The Heartwatch database contains a wealth of data which permits both cross-sectional and longitudinal analysis. It constitutes a large database, implemented and collected in a general practice setting, which indicates what is achievable in this respect. The post-cleaned data is of a high quality and allows national and health board level analysis to a high level of statistical reliability” (The National Heartwatch Programme, 2004).

## Discussion and Conclusions

*[Understanding] error provides a critical component in judging fitness for use (Chrisman 1991).*

The Independent National Data Centre (INDC) receives data from the participating practices and is responsible for data management and report production.

As a result of the data quality review, the INDC system now features full automation of the data processing to ensure it meets the agreed data quality targets; it features online facilities for both participating practices and the central administration to upload, check and correct data in addition to running financial reports and pre-defined and customized GP, regional and national demographic and clinical reports. One of most innovative features is online access to practices to their own data compared to their regional and national data (The National Heartwatch Programme, 2006).

Differences in file structures and variable naming conventions within different software systems utilised at local level are often not malleable, but once known and documented can be addressed through the creation of a common dictionary prior to merging into the central database as occurred here.

Ensuring an appreciation among participating health practitioners as to why data must be recorded in a particular manner and how it will be utilised in addition to training on how to do so is crucial. The experience and learning from the managed care and ICT dynamics of this programme will benefit practices greatly, in terms of future structured care programmes and ICT oriented initiatives. As noted internationally, there is substantial potential in capitalising on the economy of scale benefits to establish other healthcare programmes and projects which also necessitate reliable and valid data capture from general practice (Brett et al., 2006).

It has been shown that such activities can influence policy-making and planning processes through strengthening the foundation of evidence (Pirkis et al., 2006).

Some of the data quality issues identified resulted because the concept of evaluation had not been fully taken on board at the commencement of the programme, and a data quality management process was not instituted from the outset. However, once the data quality issues were identified, they were addressed and improvements implemented. Strategies for managing and improving data quality were developed and the system design and protocols enhanced. Such experience and learning about data quality from a large sector of the health care service in Ireland will facilitate further data gathering activities in this sector and throughout the healthcare environment.

## References

1. Brett, T., McGuire, S., Meade, B. and Leahy, J. (2006) "Secondary prevention of cardiovascular disease: a possible model for Australian general practice", *Australian Family Physician*, 35 (3) 157-159.
2. Capita Consulting (2005) *Evaluation of Heartwatch - Final Report: a report from Capita Consulting, commissioned by the Department of Health and Children*, unpublished.
3. Chrisman, N.R. (1991) "The Error Component in Spatial Data" In: Chapman, A.D. (2005) *Principles of Data Quality*, version 1.0., Global Biodiversity Information Facility, Copenhagen.
4. DeMaio, A. (2002) *Understanding data quality issues: finding data inaccuracies* [Online], [Retrieved January 4th 2010]. Available: [www.dama-ncr.org/Library/2002-11-12DeMaio-DataQualityIssues.ppt](http://www.dama-ncr.org/Library/2002-11-12DeMaio-DataQualityIssues.ppt)
5. Juran, J.M. (1964) "Managerial Breakthrough" In: Chapman, A.D. (2005) *Principles of Data Quality*, version 1.0., Global

- Biodiversity Information Facility, Copenhagen.
6. Lambe, B. and Collins, C. (2010) "A qualitative study of lifestyle counselling in general practice in Ireland", *Family practice*, 27 (2) 219-223.
7. Madnick, S. and Wang, R.Y. (1992) Introduction to Total Data Quality Management (TDQM) research program. TDQM-92-01. [Online], MIT Sloan School of Management, [Retrieved January 4th 2010] Available: <http://web.mit.edu/tdqm/papers/92/92-01.html>.
8. Madnick, S.E., Wang, R.Y., Lee, Y.W. and Zhu, W. (2009) "Overview and Framework for Data and Information Quality Research" *ACM Journal of Data and Information Quality*, 1 (1).
9. McGrath, E.R., Glynn, L.G., Murphy, A.W., O Conghaile, A., Canavan, M., Reid, C., Moloney, B. and O'Donnell, M.J. (2012) "Preventing cardiovascular disease in primary care: role of a national risk factor management program" *American Heart Journal*, 163 (4) 714-719.
10. Pirkis, J.E., Blashki, G.A., Murphy, A.W., Hickie, I.B. and Ciechomski, L. (2006) "The contribution of general practice based research to the development of national policy: case studies from Ireland and Australia" *Australia and New Zealand health policy*, 3: 4.
11. Redman, T.C. (2001) "Data Quality: The Field Guide." In: Chapman, A.D. (2005) *Principles of Data Quality*, version 1.0., Global Biodiversity Information Facility, Copenhagen.
12. The National Heartwatch Programme (2004) Heartwatch clinical report: March 2003 to April 2004, The Heartwatch National Programme Centre (NPC) and the Independent National Data Centre (INDC), Dublin.
13. The National Heartwatch Programme (2006) Heartwatch clinical report: March 2003 to December 2005 - Second Report, The Heartwatch National Programme Centre (NPC) and the Independent National Data Centre (INDC), Dublin.
14. US Census Bureau (2006) Definition of data quality: Census Bureau Principle Version 1.3, United States Census Bureau.