

Development of a customised data management system for a COVID-19-adapted colorectal cancer pathway

Janice Miller ^{1,2} Frances Gunn,¹ Malcolm G Dunlop,^{1,3} Farhat VN Din,^{1,3} Yasuko Maeda^{1,3}

To cite: Miller J, Gunn F, Dunlop MG, *et al*. Development of a customised data management system for a COVID-19-adapted colorectal cancer pathway. *BMJ Health Care Inform* 2021;**28**:e100307. doi:10.1136/bmjhci-2020-100307

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2020-100307>).

Received 17 December 2020
Accepted 17 June 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Colorectal Surgery, Western General Hospital, Edinburgh, UK

²Clinical Surgery, University of Edinburgh Western General Hospital, Edinburgh, UK

³Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Edinburgh, UK

Correspondence to

Miss Yasuko Maeda;
Yasuko.maeda@ed.ac.uk

ABSTRACT

Objectives A customised data management system was required for a rapidly implemented COVID-19-adapted colorectal cancer pathway in order to mitigate the risks of delayed and missed diagnoses during the pandemic. We assessed its performance and robustness.

Methods A system was developed using Microsoft Excel (2007) to retain the spreadsheets' intuitiveness of direct data entry. Visual Basic for Applications (VBA) was used to construct a user-friendly interface to enhance efficiency of data entry and segregate the data for operational tasks.

Results Large data segregation was possible using VBA macros. Data validation and conditional formatting minimised data entry errors. Computation by the COUNT function facilitated live data monitoring.

Conclusion It is possible to rapidly implement a makeshift database system with clinicians' regular input. Large-volume data management using a spreadsheet system is possible with appropriate data definition and VBA-programmed data segregation. The described concept is applicable to any data management system construction requiring speed and flexibility in a resource-limited situation.

INTRODUCTION

We introduced a COVID-19-adapted colorectal cancer pathway in an attempt to mitigate the risks of delayed and missed colorectal cancer diagnosis during the pandemic.¹⁻³ A sound information management system was required to process the flow of patients and monitor outcomes while operating this alternative pathway. As it was challenging to embed it in an existing health informatic system,⁴ we developed a bespoke data management system using Microsoft Excel and Visual Basic for Application (VBA).

METHODS

The pathway involved several different investigations with patients being signposted to the appropriate tests in a stepwise fashion depending on referral information and initial test results (figure 1).

Both administrative and clinical information was considered to estimate the information flow and risks associated with data handling. Estimates of information volume included the amount of patients to be entered into the pathway, the number of nodal information and classification required to signpost patients into the correct pathway arm as well as data needed to make live data monitoring, particularly relating to cardinal outcomes of pathway performance (eg, cancer detection rate).

Key data items required for pathway management were defined clearly. Among the two personal identifications (ID) (CHI: Community Health Index Number and UHPI: system allocated unique reference number) in the existing electronic hospital record system, UHPI was designated as the primary key ID as it was a nine-digit number that ends with a letter compared with a 12-digit numerical CHI, which provided a better property as string information.

Pathway-specific key turning points (type, combination and results of investigations) were recorded as nodal data type so that patients and/or tests could be grouped into lists or put in hierarchical order when required.

We developed a system using Microsoft Excel (2007) to maximise the functionality of the spreadsheet and incorporate VBA to minimise its limitations by segmentation of data. We kept the master spreadsheet (parent) as a .xlsx file to prevent receiving malicious VBA codes and limited the use of VBA to child files. Data were itemised, and drop-down menus were used to take advantage of the data validation function to limit entry errors. A UserForm was constructed by VBA as a user-friendly interface, to enhance efficiency of data entry and provide an overview of results. Formulae were used to allow derived computation and facilitate live monitoring of the

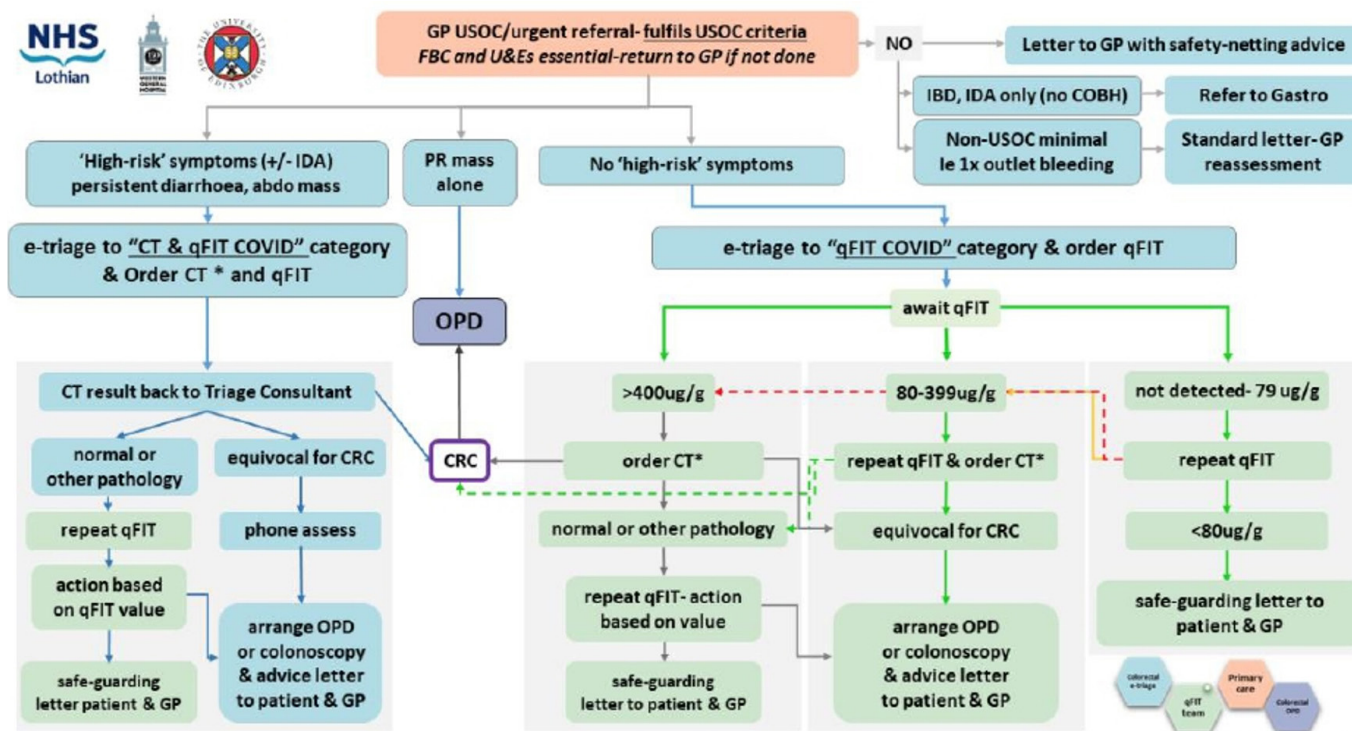


Figure 1 NHS Lothian Covid-adapted triage pathway. *order CT chest/abdomen/pelvis. CT: computed tomography scan, IDA: iron deficiency anaemia, IBD: inflammatory bowel disease, COBH: change of bowel habit, qFIT: quantitative faecal immunochemical test, OPD: outpatients department, USOC: urgent suspected of cancer.

pathway activities. Data were segregated for operational tasks with built-in loop-back data entry. The flow of information using VBA was controlled via one-way data traffic to avoid overwriting existing data in the parent file.

RESULTS

A master spreadsheet consisted of 36 items including demographical information (eg, ID, age and sex) and specific data for the pathway operation (tests, results and tasks).

Several mechanisms were introduced to minimise errors from data entry.

A formula:

=MATCH (CELL WITH UHPI)*UNC path or a mapped drive*\$label of column with UHPI]\$label of column with UHPI,0))

was used to find the matching UHPI in the main spreadsheet in order to prevent duplicate entry to the master file when a new request for a test came in. Double-check mechanism was used with 'conditional formatting' highlighting any duplicate cell.

Entry using free text was limited to two data items. All other information was divided into elemental items that could be recorded in categorical entries. Data validation was used to restrict entries to a certain data type (eg, time period to avoid wrong year). Furthermore, cell format and Text to Columns function were used periodically to apply the correct format.

In addition, any item that could be linked to unique IDs (CHI and UHPI) were automated by using the formula. Patient's sex could be worked out from the second to the last digit of the CHI, even numbers corresponding to women and odds to men. The following formula was used to avoid errors once the CHI was copied in from the electronic health record

=IF(MOD(LEFT(RIGHT(CELL,2)),2)=0, "even" , "odd") = "even" , "F" , "M")

A UserForm consisted of a function to extract data for display, update information of existing patents and add new patients with triaging information (online supplemental figure 1). This was based on VBA to display the patient's information and results of tests. The 'update' button allowed the user to overwrite information in a row with the matching UHPI in the master spreadsheet. The button 'new patient' was programmed to enter data in the next empty row after the last existing row. Entry date was regulated by the use of VBA property CDATE.

VBA was used to filter data according to different purposes. A child file ('all_patients_on_pathway.xlsx') was created using VBA macros to filter all patients included in the pathway and acted as the 'parent' files to source task lists. Advanced Filter VBA was used due to its speed and function to filter by column values with the use of AND or OR logic, to select different sets of patients for specific tasks (eg, chase results).

In a superuser file ‘tracker.xlsm’, calculations were performed using a number of formulae, mainly the COUNTIF and COUNTIFS function. These built-in formulae allowed monitoring of both performance and figures as a one-screen dashboard in real time.

An overview of the data management is summarised in online supplemental figure 2. Patient flow is summarised in online supplemental figure 3 and the pathway dashboard in online supplemental figure 4.

DISCUSSION

Development of the current data management system was born out of the need to rapidly set up a pathway for patients referred with symptoms of colorectal cancer during the COVID-19 pandemic. Modification of any existing hospital electronic systems is challenging at the best of times in public healthcare. The concept and design of this system could be used in any resource-constrained setting when the rapid development and introduction of a data management system are required. Although spreadsheets such as Microsoft Excel or Google Sheets are known to be inefficient for handling large volumes of data,⁴ they have inherent intuitiveness of direct data entry and allow computation with capabilities for ad hoc analysis.

Collection of accurate information was the key to robust pathway operations. Considerable efforts and attention to detail were required to apply multiple manual checks to run day-to-day tasks safely. Data cleaning was required also that was expected due to the makeshift nature of the system and the inherent limitations of a spreadsheet being able to function as a bona fide database. Presentation of real-time data and feedback from other clinicians was useful to develop the pathway and make necessary adjustments to data definition and validation.⁵ The flexibility and user-friendliness of the spreadsheet allowed staff to familiarise themselves with the system in a short period of time without much training. The system was easily adaptable with the rapid changes triggered during the pandemic.

Although dealing with a pandemic is a situation unique to most, managing large volumes of data using a spreadsheet is quite common both in clinical practice and in research. The described concept is applicable to any data management system construction requiring speed and flexibility.

CONCLUSION

Large-volume data management using a spreadsheet system is possible with appropriate data definition and VBA-macroprogrammed data segregation. Clinicians’ input and continuous optimisation made the system adaptable and suitable for bespoke data management in a resource-limited setting.

Acknowledgements The authors would like to thank Neil Gunn for their review of technical details.

Contributors YM designed and ran the spreadsheet. YM, JM and FG developed and refined the system and assessed its practical application. YM, MGD and FVND contributed to performance analysis and data interpretation. YM and JM wrote the manuscript. All authors critically appraised and approved the final version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Availability of data and material: A data cleaned file is available as a reference. Code availability: Available if contacted.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID ID

Janice Miller <http://orcid.org/0000-0002-0119-7736>

REFERENCES

- 1 Maeda Y, Dunlop MG, Din FVN. Risk mitigation for suspected colorectal cancer diagnostic pathway during COVID-19 pandemic. *Br J Surg*. In Press 2020;107:e361–2.
- 2 Pellino G, Vaizey CJ, Maeda Y. European Society of Coloproctology (ESCP) guideline Committee. The COVID-19 pandemic: considerations for resuming normal colorectal services. *Colorectal Dis* 2020;22:1006–14.
- 3 Santoro GA, Grossi U, Murad-Regadas S, *et al*. Delayed colorectal cancer care during COVID-19 pandemic (DECOR-19): global perspective from an international survey. *Surgery* 2021;169:30778–9.
- 4 Reeves JJ, Hollandsworth HM, Torriani FJ, *et al*. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. *J Am Med Inform Assoc* 2020;27:853–9.
- 5 Ismail L, Materwala H, Karduck AP, *et al*. Requirements of health data management systems for biomedical care and research: Scoping review. *J Med Internet Res* 2020;22:e17508.