

Using macros in microsoft excel to facilitate cleaning of research data

Justin Bauzon ^a, Caleb Murphy ^b and Sandhya Wahi-Gururaj^b

^aSchool of Medicine, University of Nevada Las Vegas (UNLV), Las Vegas, Nevada, USA; ^bDepartment of Medicine, UNLV School of Medicine, Las Vegas, Nevada, USA

ABSTRACT

Background: Retrospective chart review studies may be delayed by inability to export clean clinical data from an electronic medical record (EMR) or data repository. Macros are pre-programmed procedures that can be used in Microsoft Excel to help streamline the process of cleaning clinical datasets.

Objectives: To demonstrate how macros may be useful for researchers at community hospitals and smaller academic health centers that lack informatics support.

Methods: Using an intrinsic function of our institution's EMR, vital signs and lab results from 20 individual hospitalizations were exported to a spreadsheet. Two macros were developed to sort through these datasets and output them into a specified format. The speed of macro-assisted data cleaning was compared to manual transcription.

Results: Time spent on data cleaning was significantly reduced when using macro-assisted sorting compared to the manual approach for both vital signs (46.5 seconds versus 12.3 minutes per record, a 94% reduction; $P < 0.001$) and labs (13.7 seconds versus 2.6 minutes per record, a 91% reduction; $P < 0.001$).

Conclusions: Macros offer a flexible and efficient tool for cleaning large sets of clinical data, particularly when an institution lacks informatics support or EMR functionality to export clinical data in an analysis-ready format.

ARTICLE HISTORY

Received 12 April 2021

Accepted 7 July 2021

KEYWORDS

VBA macro; microsoft excel; retrospective chart review

1. Introduction

Retrospective chart review research depends on obtaining clinical data in an efficient manner and in a format appropriate for statistical analysis. Researchers at community hospitals and smaller academic institutions are less likely to have assistance from informatics professionals who are trained in abstracting and cleaning clinical datasets retrieved from the electronic medical record (EMR) or a central data repository. Therefore, chart review projects in these settings often rely on investigators manually abstracting data from individual patient records or having to clean large sets of raw clinical data, both of which are time-consuming processes [1].

Our team received Institutional Review Board (IRB) exemption for a chart review study that required vital sign and laboratory results from a set of prior hospitalizations. The informatics service at our community-based teaching hospital helped us identify patients meeting our inclusion criteria but was unable to provide us with the requested retrospective data. The time required to individually transcribe each vital sign and lab value from the EMR to a spreadsheet for further analysis would have caused a significant delay in our study. We identified a method of exporting the necessary data from individual patient records using a clinician-facing function of our institution's EMR, but the resulting datasets were disorganized and the process of methodically compiling this data using

standard Excel functions would have also been time-consuming.

We developed two Microsoft Excel (MS Excel) macros using Visual Basic for Applications (VBA) to facilitate the process of cleaning and compiling the vital sign and lab results exported from our EMR. Macros, short for 'macro-instruction,' are customizable software applications that transform inputs based on a pre-specified procedure or algorithm. They are a versatile tool in research and administrative settings for handling repetitive, tedious tasks in spreadsheet management and statistical analysis [2–5]. We were unable to identify any publications that explore the utility of macros for sorting and formatting raw research data. In this paper, we describe our VBA macros and compare the efficiency of macro-assisted data sorting versus manual data entry. The macros are included as an online supplement file along with instructions to provide a model for researchers at other institutions (Appendix A and B). These macros may be particularly useful for unfunded or trainee-led research projects, particularly at community hospitals and smaller academic centers where robust informatics support is not available.

2. Materials and methods

The primary study for which our team was seeking data required specific variables from patient medical

records that included vital signs (temperature, heart rate, respiratory rate, systolic and diastolic blood pressure), CBC results (leukocyte, hemoglobin, and platelet counts) and serum electrolyte panel (SEP) components (serum sodium, potassium, bicarbonate, and creatinine). To obtain this data, we used a clinician-facing function of our EMR that allows for filtering and downloading of these variables from individual patient records as a .CSV file. Two VBA macros, ‘Sort Vital Signs’ (SVS) and ‘Sort Lab Data’ (SLD), were then developed for use in MS Excel to sort the raw clinical data into a format appropriate for further analysis. We used these macros for two purposes: (1) removing data variables that were irrelevant to our study and (2) re-formatting the raw clinical data into a format conducive to statistical analysis, as the raw data included many interspersed text boxes, blank cells, rows, and columns.

To apply the SVS and SLD macro, the raw clinical data was copied from the exported file into the ‘EMR Sorter’ spreadsheet. Activating the SVS macro within MS Excel eliminated any information irrelevant to our study and sorted the vital sign data in an orderly, chronological fashion (Figure 1). Similarly, when activated, the SLD macro removed all unnecessary laboratory data and sorted the remaining relevant data (Figure 2). Detailed instructions on using the SVS and SLD macros are included in Appendix A (‘EMR Sorter Instructions’). These macros are included in Appendix B (‘EMR Sorter.xlsm’).

2.1. Performance evaluation of the VBA macros

Time required to sort the relevant data (selected vital signs and lab results) from a sample of 20 patient

charts into a spreadsheet using the VBA macro approach was compared to manual data entry using an unpaired t-test. Accuracy of data input between approaches was determined by comparing final data sets for consistency. Appropriate data security practices were followed throughout. Dates and times within figures were altered to maintain HIPAA compliance.

3. Results

The 20 sample hospitalizations included 1276 separate clinical data points, including 1059 vital signs, 108 CBC results, and 109 SEP component values. On average, there were 53 vital signs and 11 lab records (CBC and SEP) per record.

Using the manual entry approach, the vital sign data was entered into a spreadsheet in the desired format in 246.7 minutes, for an average of 738 seconds (or 12.3 minutes) per patient record (Figure 3). With the SVS macro, the vital signs data was sorted into the desired format in 15.5 minutes, for an average of 46.5 seconds per patient record ($P < 0.001$).

Manual transcription of the 20 patients’ lab results totaled 51.3 minutes, with a mean result of 154.1 seconds (or 2.6 minutes) per patient record (Figure 4). Using the SLD macro in MS Excel, lab data was transferred into the desired format in 4.6 minutes, averaging 13.7 seconds for each patient record ($P < 0.001$).

Comparison of macro-assisted versus manually sorted results showed identical datasets, suggesting excellent accuracy in the overall transfer of data using macros.

	A	B	C	D	E	F	G	H	I	J	K	L
1	ED to Hosp-Admission											
2	Discharged											
3	Friday, January 1, 2010											
4	Research Medical Center											
5	Default Flowsheet Data (all recorded)											
6	Vital Signs											
7	Raw Name	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10
8	Vital Signs											
9	Temp	1543	1300	1111.39	1033	0840						
10	Temp sec			36.7 °C (98 °F)								
11	Pulse			DE (t) MU (t) at 01:05:19 1111								
12	Resp			DE (t) MU (t) at 01:05:19 1111								
13	BP			DE (t) MU (t) at 01:05:19 1111								
14	MAP (mmHg)			126/78								
15	BP Location			DE (t) MU (t) at 01:05:19 1111								
16	Patient Position			Supine								
17	Oxygen Therapy			DE (t) MU (t) at 01:05:19 1111								
18	SpO2			91 %								
19	Oxygen Therapy			None (blow-by)								
20	Pain Assessment			None (blow-by)								
21	Pain Assessment	Noninvasive pain	Noninvasive pain	Noninvasive pain	Noninvasive pain							
22	Pain Score	KM at 01:05:19 1543	KM at 01:05:19 1443	KM at 01:05:19 1228	KM at 01:05:19 0937							
23	Pain Score	0 - No pain	0 - No pain	0 - No pain	0 - No pain							
24	Pain Score	KM at 01:05:19 1543	KM at 01:05:19 1443	KM at 01:05:19 1228	KM at 01:05:19 0937							
25	Raw Name	0734	1/5/10	712:29	0500	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10	1/5/10
26	Vital Signs											
27	Temp			36.9 °C (98 °F)								
28	Temp sec			DE (t) MU (t) at 01:05:19 0712								
29	Pulse			DE (t) MU (t) at 01:05:19 0712								
30	Pulse			77								

BEFORE



	A	B	C	D	E	F	G
1	RowName	Temp (C)	Pulse	Resp	SBP	DBP	
2	1/1/10		133	18	116	73	
3	1/1/10		128	16	115	68	
4	1/1/10	38.4					
5	1/1/10		136	18	103	59	
6	1/1/10	37.5	134	17	109	62	
7	1/2/10	37.6	112	17			
8	1/2/10		109	112	71		
9	1/2/10	38	111	18	103	65	
10	1/2/10	38	111				
11	1/2/10	37.9	100	18			
12	1/2/10		96				
13	1/2/10	36.9	88	18			
14	1/2/10	37.7	101				
15	1/2/10	37.8	102				
16	1/3/10	37.6	88	100	65		
17	1/3/10	39.3	114	17			
18	1/3/10	37.7	102	17	110	70	
19	1/3/10	36.6	90	17	102	68	
20	1/3/10	36.7	98	17	104	64	
21	1/3/10	38.6	102	17	120	74	
22	1/3/10	39.3					
23	1/3/10	37.4					
24	1/3/10	37.3	99	17	101	57	
25	1/3/10	36.7	85	17	110	73	
26	1/3/10	37.3	95	17	111	72	
27	1/3/10	37.3	95	17	111	72	
28	1/3/10	36.9	96	17	116	77	
29	1/3/10	36.9	91	18	111	72	
30	1/3/10	37.1	80	18	109	66	
31	1/4/10	36.9	78	18	110	71	
32	1/4/10	36.7	78	16	123	76	
33	1/4/10	36.7	86	18	115	69	
34	1/4/10	36.5	79	16	123	76	
35	1/4/10	36.8	78	17	123	76	
36	1/4/10	36.9	77	17	132	85	
37	1/4/10	37.7	114	16	103	68	
38	1/5/10	36.8	75	17	137	81	
39	1/5/10	36.9	77	18	130	84	
40	1/5/10	36.7	80	18	126	78	

AFTER

Figure 1. Comparison of exported vital signs before and after sorting.

	A	B	C	D	E
1	Component	White Blood Cell	RBC'S	Hemoglobin	
2	Labset Ref Req & Units				
3	1510	3.10 - 10.20 KMM3	4.08 - 5.70 MMMS	13.1 - 16.8 GDL	
4	1510 PM	1510			
5	1510 PM		8.7	4.31	13.1
6	1510 PM				
7	1510 PM				
8	1510 PM				
9	1510 PM				
10	1510 PM				
11	1510 PM		4.8 (5.0)	11.8 (1)	
12	1510 PM				
13	1510 PM				
14	1510 PM		7	4.37 (2.4)	
15	1510 PM				
16	1510 PM				
17	1510 PM				
18	1510 PM		4.2	4.27 (2.3)	
19	1510 PM				
20	1510 PM				
21	1510 PM				
22	Component	Hematocrit	Mean Cell Volume	Mean Cell Hemogr	
23	Labset Ref Req & Units				
24	1510	38.2 - 48.4 %	80.1 - 98.7 FL	27.2 - 34.3 Hg	
25	1510 PM				
26	1510 PM		38.4	85.1	28
27	1510 PM				
28	1510 PM				
29	1510 PM				
30	1510 PM		36.1 (1)	84.8	29.1
31	1510 PM				
32	1510 PM				
33	1510 PM				
34	1510 PM				
35	1510 PM		36.8 (1)	84.1	28.4
36	1510 PM				
37	1510 PM				
38	1510 PM				
39	1510 PM		36.8 (1)	83.5	28.3
40	1510 PM				
41	1510 PM				
42	1510 PM				
43	Component	Mean Cell Hemoglobin-Cor Platelet			
44	Labset Ref Req & Units				
45	1510	32.7 - 38.4 %	119 - 332 KMMMS		
46	1510 PM				
47	1510 PM				
48	1510 PM		34.1	202	
49	1510 PM				
50	1510 PM				

BEFORE



	A	B	C	D	E	F	G	H	I	J
1		White Blood Cell	Hemoglobin	Platelet	Sodium	Potassium	Total CO2	Creatinine		
2		3.10 - 10.20 KMM3	13.1 - 16.8 GDL	119 - 332 KMM3	136 - 145 MMOL/L	3.5 - 5.1 MMOL/L	20 - 31 MMOL/L	0.55 - 1.30 mg/dL		
3	1510 CBC		8.7	13.1	202					
4	1510 SEP				136	3.6	29	0.81		
5	1510 CBC		4.8	11.8	191					
6	1510 SEP				145	3.6	25	0.88		
7	1510 CBC		7	12.4	213					
8	1510 SEP				137	3.5	27	0.75		
9	1410 CBC		4.2	12.3	212					
10	1410 SEP				140	3.8	29	0.74		
11										

AFTER

Figure 2. Comparison of lab results before and after sorting.

4. Discussion

Electronic medical records and their use in hospital systems as repositories for clinical data collection can facilitate retrospective chart review-based research [6]. However, smaller academic centers and community hospitals may lack the necessary informatics support or software tools to export data from EMRs in a format conducive to analysis [7]. Furthermore, delays in accessing clinical data due to hospital constraints and underinvestment in informatic resources may lead to increased study costs [8].

Macros offer a flexible and perhaps underutilized tool for assisting researchers in obtaining clean, analyzable datasets. Our study demonstrates the efficiency of using

macros as a tool for cleaning research data that may not be formatted correctly for analysis. Use of macros contributed to a 90% reduction in data acquisition and sorting time compared to manually entering data into a spreadsheet. Macros follow prewritten algorithms, virtually eliminating the possibility of erroneous data entry by humans if coded correctly. Larger datasets and varied levels of personnel training may be prone to greater error rates [9].

Our study demonstrates that creative approaches exist for researchers to overcome barriers in obtaining retrospective clinical data when traditional avenues, such as being provided analysis-ready data by an informatics service, are not available. Furthermore, the VBA macros described in this paper (and shared online) can be

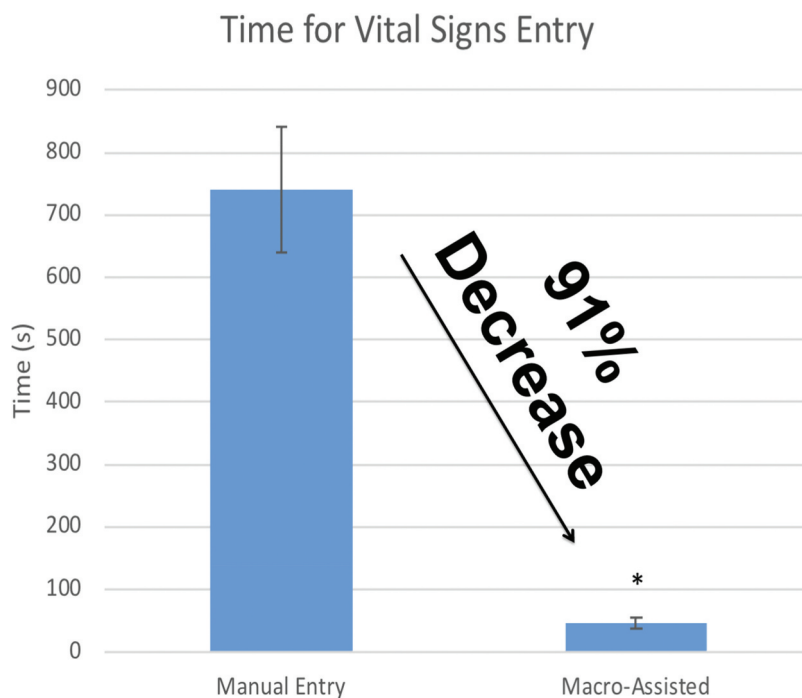


Figure 3. Time for vital signs entry.

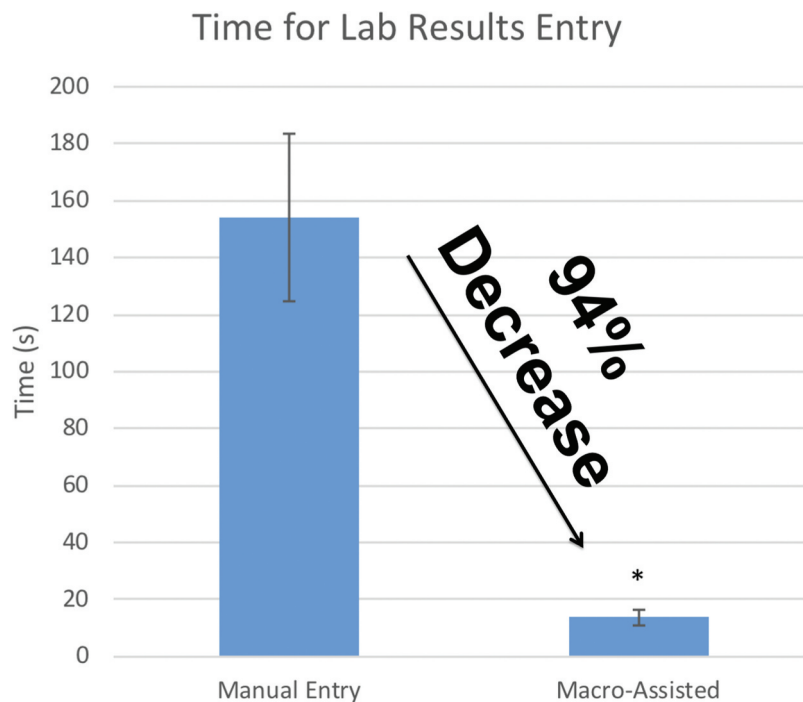


Figure 4. Time for lab results entry.

modified and repurposed for organizing a multitude of clinical data categories and formats. Access to MS Excel is available at most institutions, which allows for the facilitation of macros as a cost-friendly alternative to purchasing other commercial software for data analysis. MS Excel on its own presents its own challenges in medical research, many of which can be addressed with macro programming [10]. This is particularly relevant for medical trainees who are tasked with reviewing medical records for clinical data acquisition or organizing datasets.

There are some limitations to a macro-assisted approach to data cleaning. First, it is most useful if data can be obtained in bulk. In our study, vital sign and lab record data were exportable on a patient-by-patient basis from the Epic EMR, which allowed approximately 50 vital signs and 20 lab components to be sorted at a time. Second, the efficiency of macros may change based on the technical capabilities of the computer processors available to the user. Third, although VBA macros are versatile enough to be repurposed for other data formatting requirements, modification requires some basic understanding of programming. Completion of both macros for the purposes of this study took approximately a week, but this may vary widely depending on skill level. Finally, this approach requires careful adherence to HIPAA-compliant data security practices, although this is true for any retrospective chart review study.

Author's contributions

JB developed the macro programming used in the study, performed the manual entry portion of the analysis, and

was a major contributor in writing the manuscript. CM organized the study and analyzed the macro-processed data, as well as contributed to the manuscript. SW oversaw completion of the study and contributed to the manuscript. All authors read and approved the final manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Justin Bauzon <http://orcid.org/0000-0002-0958-4343>
Caleb Murphy <http://orcid.org/0000-0003-3162-3780>

References

- [1] Sarkar S, Seshadri D. Conducting record review studies in clinical practice. *J Clin Diagn Res.* 2014;8(9):JG01–JG4.
- [2] Deochand N, Costello MS, Fuqua RW. Recording an excel® macro to specify date ranges for clinical data. *Behav Anal Pract.* 2016;9(3):230–234.
- [3] Talay-Değirmenci I, Holmes CJ, Kuo PC, et al. An analytic decision support tool for resident allocation. *J Surg Educ.* 2013;70(1):31–35.
- [4] Aliane N. Data acquisition and real-time control using spreadsheets: interfacing excel with external hardware. *ISA Trans.* 2010;49(3):264–269.
- [5] Khan HASCEW, Microsoft A. Excel add-in for easy creation of survival curves. *Comput Methods Programs Biomed.* 2006;83(1):12–17.
- [6] De Moor G, Sundgren M, Karla D, et al. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform.* 2015;53:162–173.

- [7] Takenouchi K, Yuasa K, Shioya M, et al. Development of a new seamless data stream from EMR to EDC system using SS-MIX2 standards applied for observational research in diabetes mellitus. *Learn Health Syst.* 2018;3(1):e10072.
- [8] Barick U, Vijaykanth A, Bharucha H, et al. are retrospective patient chart audits an affordable and reliable answer to healthcare data needs? Assessing the ground reality. *Biomed J Sci Tech Res.* 2018;7(2). DOI:10.26717/BJSTR2018.07.001476
- [9] Worster A, Haines T. Advanced statistics: understanding Medical Record Review (MRR) studies. *Acad Emerg Med.* 2004;11(2):187–192.
- [10] Tanavalee C, Luksanapruksa P, Singhatanadgige W. Limitations of using microsoft excel version 2016 (MS Excel 2016) for statistical analysis for medical research. *Clin Spine Surg.* 2016 Jun;29(5):203–204. .

Appendices

Appendix A: EMR_Sorter_Instructions.docx

Appendix B: EMR_Sorter.xlsm

EMR Sorter Instructions

- Enabling Macros**(1) Open the 'EMR Sorter.xlsm' workbook.
- (2) A yellow 'Security Warning' bar will appear when you first open a workbook containing macros. Select 'Enable Content' → Select Enable Macros from the window that pops up.
- a. Note: only enable macros from a secure or trusted source.

Sort Vital Signs (SVS) Macro

- (1) Export the desired vital sign data into a spreadsheet format (.XLS, .CVS).
 - a. See 'Sample Vitals' tab for example of exported vital sign data.
- (2) Select all of the vital sign data in the spreadsheet → Copy the data.
- (3) In the 'Raw Data' tab in the workbook, paste the data into cell A1 (by default, the cell will contain the text 'COPY/PASTEHERE').
- (4) Click the 'Sort Vital Signs (SVS Macro)' button on the spreadsheet.
- (5) Enter a patient identifier in the popup window → select OK. The macro will start immediately.
 - a. Depending on how much data is present, this may take some time.
- (6) The data will be exported in the desired format onto the 'Raw Data' tab once the macro has completed.

Sort Lab Data (SLD) Macro

- (1) Export the desired lab data into a spreadsheet format (.XLS, .CVS).
 - a. See 'Sample Labs' tab for example of exported vital sign data.
 - (2) Select all of the lab data in the spreadsheet → Copy the data.
 - (3) In the 'Raw Data' tab in the workbook, paste the data into cell A1 (by default, the cell will contain the text 'COPY/PASTEHERE').
 - (4) Click the 'Sort Lab Data (SLD Macro)' button on the spreadsheet.
 - (5) Enter a patient identifier in the popup window → select OK. The macro will start immediately.
 - a. Depending on how much data is present, this may take some time.
 - (6) The data will be exported in the desired format onto the 'Raw Data' tab once the macro has completed.
- **Please do not hesitate to contact the authors for any clarifications.