



How to enter data with Excel in health research



Vo Tuan Khoa



Epidemiological Research Training Course VI-2, 2015

Outline

- Introduction to Excel
- Prepare data with Excel
- Data enter with Excel

Introduction to Excel

- Microsoft Excel is a useful spreadsheet program
- Not designed to be a research data entry tool → commonly used because almost researcher knows how to use it basically
- Data files from Excel can be shared or imported directly into files formatted by most statistical software (SPSS, Stata, R, Minitab)
- Limitation: spreadsheets containing less than 256 variables (columns) and 65,536 records (rows).

Column label (variable)

Cell (data)

Row (record)

	A	B	C	D	E	F	G	H	I
1	doe	subid	yob	job	edu	height	weight	waist	smoking
2	21/2/2012	1	1961	2	3	153	55	80	1
3	22/2/2012	2	1952	2	2	164	55	74	1
4	23/2/2012	3	1969	1	4	157	89	111	2
5	24/2/2012	4	1955	3	2	170	58	85	3
6	25/2/2012	5	1949	1	4	162	52	78	1
7	26/2/2012	6	1953	2	3	162	62	91	2
8	27/2/2012	7	1953	1	3	160	70	99	2
9	28/2/2012	8	1964	1	4	179	73	90	1
10	29/2/2012	9	1955	1	4	158	57	88	1
11	1/3/2012	10	1949	3	3	164	77	94.5	1
12	2/3/2012	11	1948	3	3	164	59	90	2

Prepare data with Excel

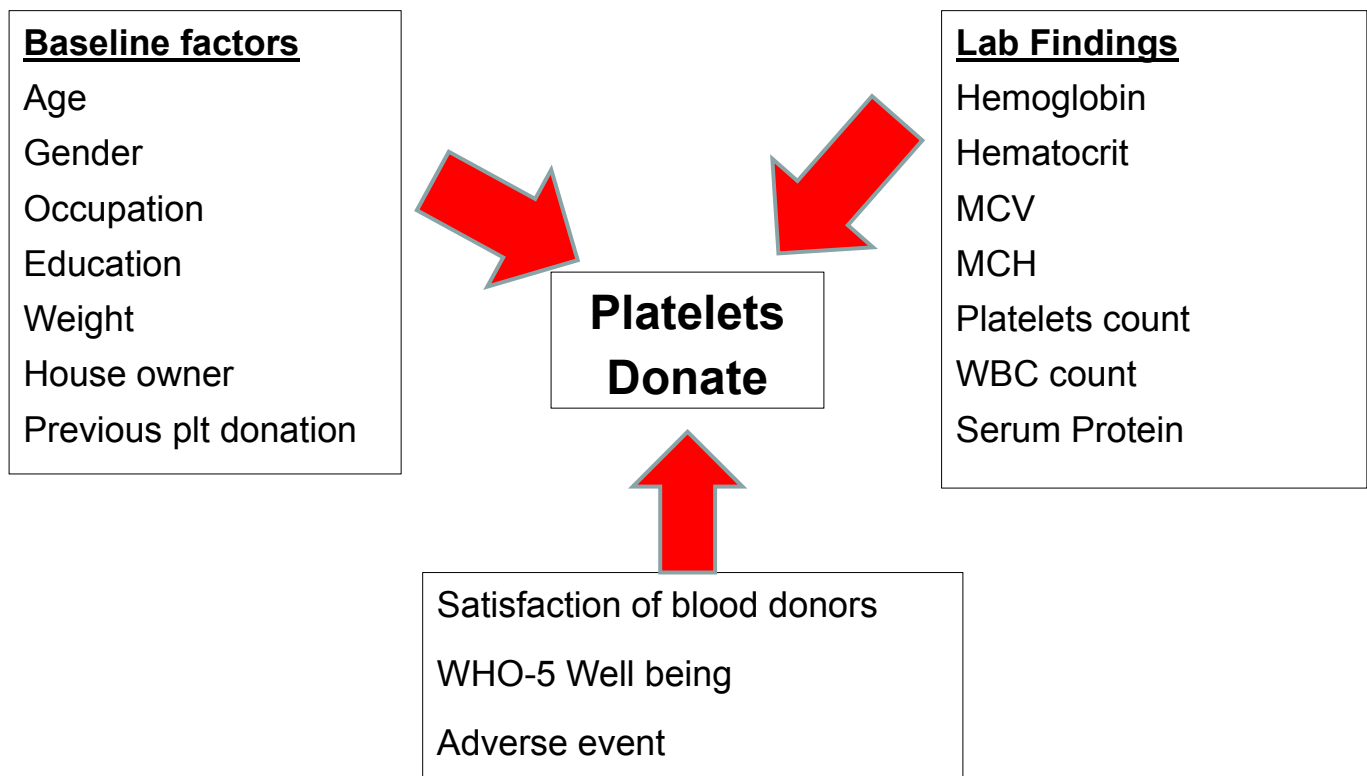
1. Well-designed protocol (conceptual framework)
2. Questionnaire form (verify all variables if necessary and clarify format)
3. Making data dictionaries
4. Making data tables

PROTOCOL

Group of Hematology

**repeat Double platelet apheresis for
donation to Compare the safety
duration: an interventional study in
vietnam
(DRACULA study)**

CONCEPTUAL FRAMEWORK



Phụ lục 9



BỆNH VIỆN TRUYỀN MÁU HUYẾT HỌC

Nghiên cứu BTH-TC2014

MSĐT nghiên cứu: □/□□/□□□

Ngày thu thập: □□ - □□ - 2015

Lần hiến: Đầu tiên

Người thu thập:

Số hồ sơ:

THÔNG TIN CĂN BẢN BAN ĐẦU

- Năm sinh: 19 □□
- Phái tính:
 - Nam
 - Nữ
- Số lần đã hiến tiểu cầu: □ □
- Trình độ học vấn:
 - Biết đọc biết viết hoặc tiểu học
 - Trung học cơ sở
 - Trung học phổ thông
 - Cao đẳng / Đại học hoặc trên đại học



BỆNH VIỆN TRUYỀN MÁU HUYẾT HỌC

Nghiên cứu BTH-TC2014

MSĐT nghiên cứu: □/□□/□□□

Ngày thu thập: □□ - □□ - 2015

Lần hiến: □□

Người thu thập:

Số hồ sơ:

PHIẾU THEO DÕI HIẾN TIỂU CẦU

(Dành cho nhân viên thực hiện chiết tách tiểu cầu thu thập)

I. Các chỉ số xét nghiệm trước và sau khi hoàn thành chiết tách tiểu cầu:

Chỉ số	Trước chiết tách	Sau chiết tách	Túi tiểu cầu
1. Bạch cầu ($10^3/\text{mm}^3$)	□ □.□	□ □.□	
2. Hồng cầu ($10^6/\text{mm}^3$)	□.□□	□.□□	

III. Các phản ứng lâm sàng không mong muốn xảy ra trong quá trình chiết tách :**A. Những phản ứng ở mức độ nhẹ:**

- | | | |
|----------------------------|-------|----------|
| 1. Tê môi, tê tay, tê chân | 1. Có | 2. Không |
| 2. Khó chịu | 1. Có | 2. Không |
| 3. Chóng mặt | 1. Có | 2. Không |
| 4. Nhức đầu | 1. Có | 2. Không |
| 5. Vã mồ hôi | 1. Có | 2. Không |
| 6. Xanh xao | 1. Có | 2. Không |
| 7. Xiu (bất tỉnh) | 1. Có | 2. Không |

B. Những phản ứng ở mức độ trung bình:

- | | | |
|-------------|-------|----------|
| 1. Buồn nôn | 1. Có | 2. Không |
| 2. Nôn | 1. Có | 2. Không |



BỆNH VIỆN TRUYỀN MÁU HUYẾT HỌC

Ngày thu thập: □□ - □□ - 2015

Nghiên cứu BTH-TC2014

Lần đến: Đầu tiên

MSĐT nghiên cứu: □/□□/□□□

Số hồ sơ:

CÂU HỎI ĐÁNH GIÁ VỀ VIỆC HIẾN MÁU TÌNH NGUYỆN

(Dành riêng cho người hiến tiểu cầu)

Hướng dẫn trả lời câu hỏi:

Hãy trả lời từng câu bằng cách khoanh tròn vào số tương ứng với câu trả lời. Không có câu trả lời nào đúng hay sai. Nếu chưa chắc chắn, vui lòng chọn câu trả lời mà bạn cho là phù hợp với bạn nhất.

1. Đánh giá chung của bạn về hiến máu

	Có	Không	Không rõ
--	----	-------	----------

Data dictionaries and/or code book

- Data dictionary makes the column definitions explicit
- Data dictionary is a table of information about the database itself
 - rows representing fields
 - column for field name, field type and field description

viết tắt	tên biến	loại biến	giá trị	ghi chú
subid	mã số	chuỗi	xxxxxx	
ho	họ và chữ lót	chuỗi		không dấu
ten	tên	chuỗi		không dấu
ngaync	ngày vào nghiên cứu	ngày	nn/tt/nnnn	
nhom	nhóm ngẫu nhiên	phân loại	2.nhóm 2 tuần	
			3.nhóm 3 tuần	
			4.nhóm 4 tuần	
ngaysinh	ngày sinh	liên tục	xxxx	1997-1945
phai	phái tính	nhị giá	1.nam	
			2.nữ	
hientc	số lần hiến tiểu cầu	liên tục		
hocvan	trình độ học vấn	phân loại	1.cấp I	
			2.cấp II	
			3.cấp III	
			4.đại học/cao đẳng	
vieclam	công việc làm	phân loại	1.trí óc	
			2.chân tay	
			3.hưu trí	
			4.không	

Variable names

- Most statistical programs allow long column headings or variable names
- Some rules for variable names
 - short enough to type quickly but long enough to be descriptive
 - English meaning
 - avoiding spaces and special characters (especially “dấu tiếng Việt”)

Coded responses vs Free text

- Defining a variable should include specifying its range of allowed values
- Limiting responses to a ranged coded value > allowing free-text responses
- Set of response options to a question
 - **exhaustive** (all possible options are provided)
 - **mutually exclusive** (no two options are both correct)
- Consistent for coding yes/no (dichotomous) variables
- Consider “All that apply” questions

Data tables

- All computer databases have one or more data tables
 - rows = records or **entities**
 - columns = fields or **attributes**
- Simplified data table
 - each row = an individual subject
 - each column = a subject-specific attribute (name, age, sex, predictor and outcome variable)
- Should assign a unique **identification number** (subject ID) to each study participant

Microsoft Excel - Data tables

File Home Insert Page Layout Formulas Data Review View Add-Ins

Normal Page Layout Page Break Preview Custom Views Full Screen

Workbook Views

Ruler Formula Bar Gridlines Headings Show

Zoom 100% Zoom to Selection Zoom

New Window Arrange All Freeze Panes Unhide Window

View Side by Synchronous Reset Window

	A	B	C	D	E	F	G	H	I	J
1							1.nam		1.cap II	1.tri oc
2					2.moi 2 tuan		2.nu		2.cap II	2.chan tay
3					3.moi 3 tuan				3. cap III	3.huu
4					4.moi 4 tuan			(so lan)	4.dh	4.khong
5	subid	ho	ten	ngaync	nhom	namsinh	phai	hientc	hocvan	vieclam
6										
7										
8										
9										

Data entry with Excel

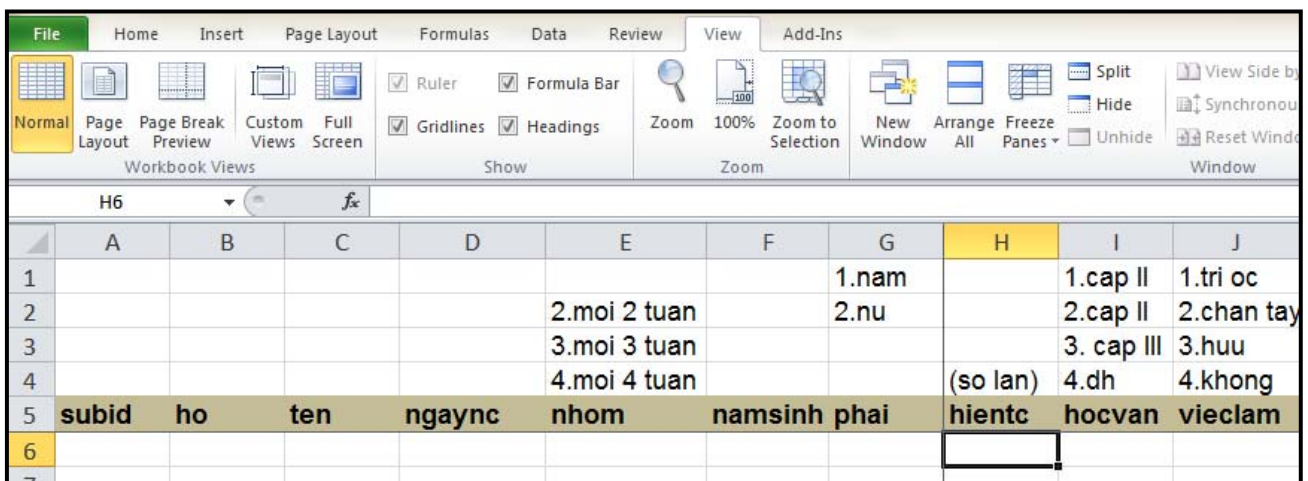
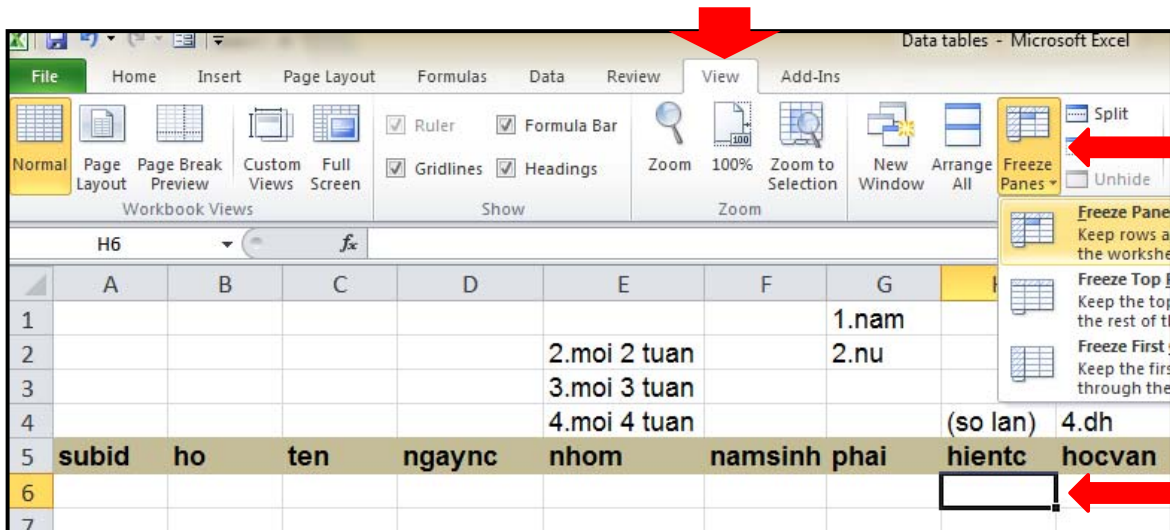
- Two in one (one reads code, one enters data)
 - Tab
 - Freeze panes tool
 - Data validation tool
- Data form: each subject for each data entry form

Freeze panes

To view only and retain some top rows (such as row 1 contains your variable names) and some important column (such as column 1 contains study id) when scroll bars move

- Put your cursor in the cell that is simultaneously below the rows you want to freeze and to the right of the columns you want to freeze
- View / Freeze Panes: select Freeze Panes and undo by selecting Unfreeze Panes

MS Excel for Public Health



Data validation tool

- To limit value in a column to a certain range or a set of values
- To prevent invalid value from being entered into a cell

Data validation tool

Highlight the column you want and then select Data / Data validation

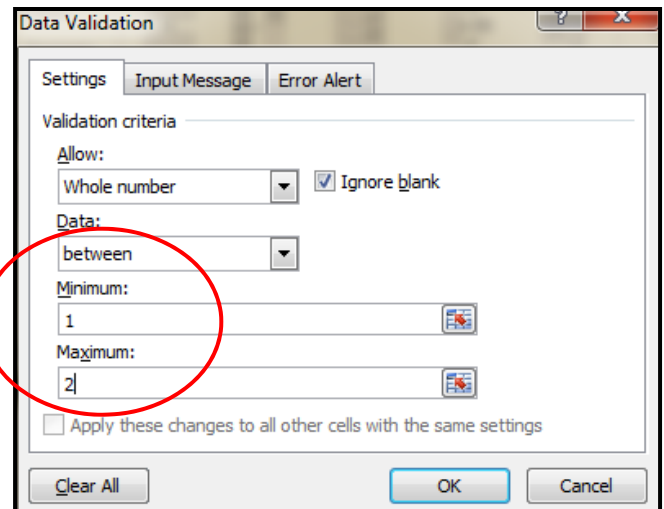
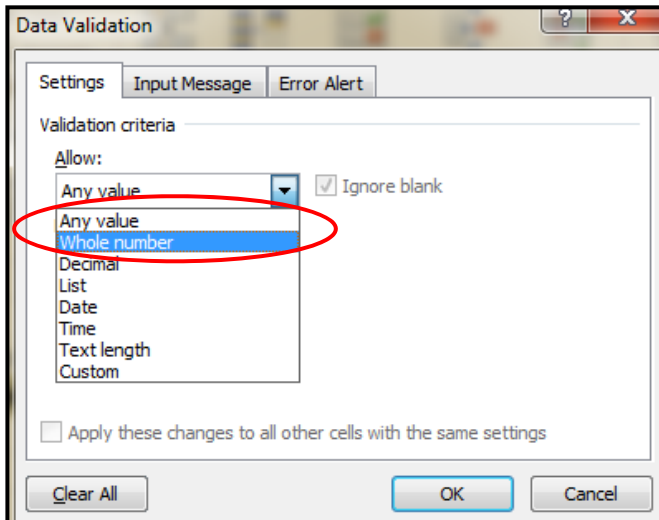
The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The 'Data Validation' button in the ribbon is highlighted with a red arrow. Another red arrow points to column G in the spreadsheet, which is highlighted in blue. A third red arrow points to the 'Data Validation' dropdown menu. The spreadsheet contains data for a form, with column G labeled 'phai' and containing the value '1'.

ten	ngaync	nhom	namsinh	phai	hientc	hocvan
Van	1/7/2015	2	1972	1		2
Thi	2/7/2015	3	1990			

Data Validation
Prevent invalid data from being entered into a cell.
For example, you could reject invalid dates or numbers greater than 1000.
You can also force input to be chosen from a dropdown list of values you specify.
Press F1 for more help.

Data validation tool

In pop up menu: select type of data you want to enter and specify the range



Data validation tool

Now, we enter an invalid value, an error message pops up

subid	ho	ten	ngaync	nhom	namsinh	phai	hientc	hocvan	vieclam
1	Nguyen	Van	1/7/2015	2	1972	1		2	1
2	Le	Thi	2/7/2015	3	1990	3			

An error message dialog box from Microsoft Excel is shown, stating: "The value you entered is not valid. A user has restricted values that can be entered into this cell." The 'Retry' button is circled in red. A red arrow points to the value '3' in the 'phai' column of the second row in the table above.

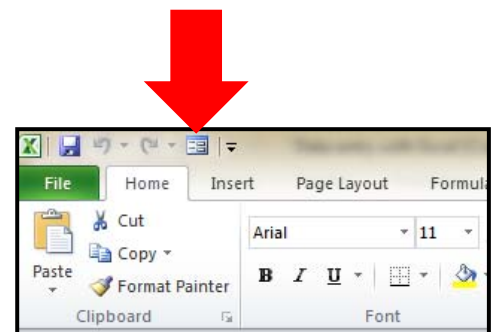
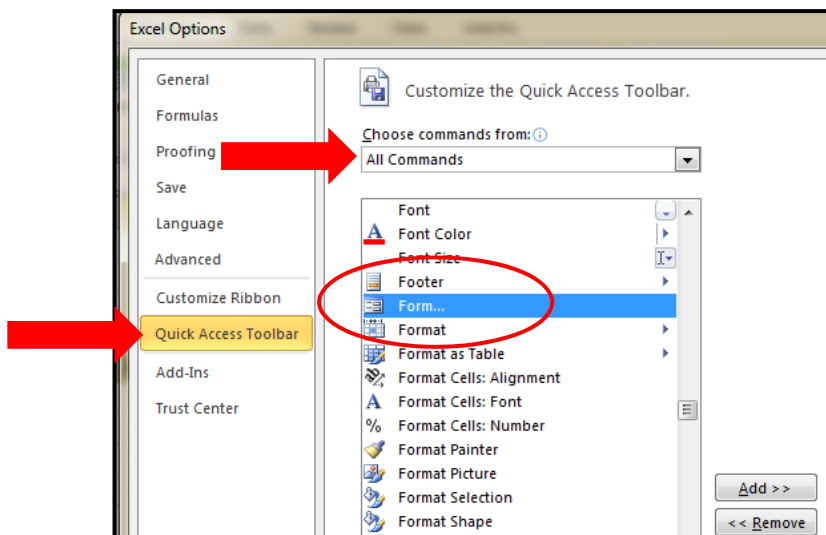
Note: **not detect** the incorrect value if it lies in the specified range

Data form tool

Create a form in excel, and when enter the form, data will be entered into the spread sheet

Data form tool

- File / Options
- In pop up menu: select *Quick Access Toolbar*, *All Commands* and then *Form*, *Add* and *OK*



Data form tool

- Highlight all columns you want
- Click icon Form in corner of left top
- Then enter data in data form

subid	ho	ten	ngaync	nhom	namsinh	phai	hientc	hocvan	viclam	thuocla	ruoubia	nha
					1.nam			1.cap II	1.tri oc	1.co	1.nhieu	1.co
				2.moi 2 tuan	2.nu			2.cap II	2.chan tay	2.khong	2.it	2.khong
				3.moi 3 tuan				3. cap III	3.huu		3.khong	
				4.moi 4 tuan			(so lan)	4.dh	4.khong			(so huu)

Sheet2

subid:

hg:

ten:

ngaync:

nhom:

namsinh:

phai:

hientc:

hocvan:

viclam:

thuoda:

ruoubia:

nha:

New Record

New

Delete

Restore

Find Prev

Find Next

Criteria

Close

References

1. Alan C. Elliott, Linda S. Hynan, Joan S. Reisch, Janet P. Smith. 2006. Preparing Data for Analysis Using Microsoft Excel. *Journal of Investigative Medicine*. 54(6); 334-342.
2. Microsoft Excel for data entry – Health research. 2010. Fernandez Hospital, Hyderabad (<https://fernandezresearch.files.wordpress.com/2010/09/data-entry-with-excel.pdf>)