

Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability & Accountability Act (HIPAA) Privacy Rule

Medical Research Office

Nov 2019

Brief Overview (from UNR Website):

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 was enacted by the U.S. Congress to regulate the protection of private health information for individuals. Protected Health Information (PHI) is individually identifiable health information transmitted or maintained in any form or medium (electronic, oral, or paper) by a covered entity or its business associates. For more information see: <https://www.unr.edu/research-integrity/human-research/researchers-affiliates/clinical/regulatory/hipaa-guidance>

Find the appropriate section for your Institutional Review Board (IRB) approval:

- I. No IRB approval
- II. IRB approval for no collection/storage of PHI
- III. IRB approval for collection and storage of PHI

** If your data was collected or is stored in California see their additional state-specific guidelines: <https://www.dhcs.ca.gov/dataandstats/Documents/DHCS-DDG-V2.0-120116.pdf>*

I. No IRB approval (from IRB website):

No human subject research conducted by University/Affiliate researchers (including student research) may be initiated until Research Integrity publishes written documentation of exempt determination or IRB approval. The jurisdiction the IRB covers all research involving human subjects conducted at the University of Nevada, Reno, the University School of Medicine, and Affiliates (i.e. Desert Research Institution, Truckee Meadows Community College, Saint Mary's Regional Medical Center, VA Sierra Nevada Health Care System, Nevada Department of Health & Human Services, and Renown Health). For more information and to apply for IRB review please visit: <https://www.unr.edu/research-integrity/human-research>

II. IRB approval for no collection/storage of PHI

If the approval you obtained from IRB to accumulate your data does not cover PHI then you will need to avoid collecting any of the following types of information for the entirety of your research and thereafter until receiving written permission from and IRB to do so:

- “(A) Names
- (B) All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census:
 - (1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and

- (2) The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000
- (C) All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
- (D) Telephone numbers
- (L) Vehicle identifiers and serial numbers, including license plate numbers
- (E) Fax numbers
- (M) Device identifiers and serial numbers
- (F) Email addresses
- (N) Web Universal Resource Locators (URLs)
- (G) Social security numbers
- (O) Internet Protocol (IP) addresses
- (H) Medical record numbers
- (P) Biometric identifiers, including finger and voice prints
- (I) Health plan beneficiary numbers
- (Q) Full-face photographs and any comparable images
- (J) Account numbers
- (R) Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section [Paragraph (c) is presented below in the section “Re-identification”]; and
- (K) Certificate/license numbers”

<https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>

Please note: not having IRB approval to collect/store PHI does NOT mean that you cannot link observations over time! Your data analysis will always be stronger if you maintain all of your time series data and this can easily be done without the collection of PHI. For example, imagine you are administering a self-assessment pre and post intervention: you can link individuals to each survey, you simply cannot know or store the names (or other PHI) of the individuals. A simple way to accomplish this task is to ask participants to create their own identification code. Simply ask study participants to create an identification code/alias for themselves that:

- (1) they can easily remember the next time they are completing the survey,
- (2) that does not use/contain their name, birthdate, or other PHI, and
- (3) that cannot reasonably be linked to them in any way other than by their own memory.

Be sure to have a data field on your surveys for “identification code” and then the code can be used to link observations over time (i.e. from different survey periods). If you use this method be sure to reiterate to participants, each and every time that they complete a survey that they need to use their same, self-determined code for the “identification code” field.

III. IRB approval for collection and storage of PHI

These instructions are designed for data stored in Microsoft Excel spreadsheets. If you are using a different database and would like help de-identifying PHI please contact the medical research office. Below are de-identification instructions based on the “Safe Harbor” method from the U.S. Department of Health & Human Services which were created using fictitious PHI:

- Step 1: Create Codes for your Dataset
 - Insert a column on the far left side of your datasheet (usually column A, see below).

File

Home

Insert

Page Layout

Formulas

Data

Review

View

Acrobat

Tell me what you want to do...

Sign in

</

- Assign IDs to each individual using a unique code:
 - The code must not be derived from or related to information *about the individual* (but can be related to the treatment/exposure the individual received) and is not otherwise capable of being translated so as to identify the individual.
 - Use names that make sense on some level (usually based on restrictions to randomization) unless the data is from a simple random sample (in that case just use numbers; see below).
 - Right click on the column and select format cells:

</

- Select “text” and choose “ok”: this will allow you to use codes like “0001” and combinations of letters and numbers as you see fit.

- To save time you can “pull down” your naming scheme to the bottom of the list rather than typing it each time:

	A	B	C	D	E	F	G	H	I
1	Master Code	Medical Record	First Name	Last Name	Age	Gender	Tobacco	Total Cholesterol	Fam Hx
2	MC00001	0793073	George	Thompson	35	M	NS	211	Positive
3	MC00002	0657479	Hary	Hagen	66	M	NS	128	Positive
4	MC00003	4400769	Molly	Walton	37	F	NS	163	Positive
5	MC00004	0501127	Marie	Cobb	44	F	NS	142	Negative
6	MC00005	4370320	Bob	Day	37	M	FS	227	Positive
7	MC00006	3154883	Sally	Hahn	42	F	NS	302	Negative
8	MC00007	4522135	Christy	Kelly	33	F	NS	251	Positive
9	MC00008	3044140	Sharon	Lowe	30	F	NS	261	Positive
10	MC00009	0758076	Terri	Mitchell	37	F	NS	310	Negative
11	MC00010	0704755	Toni	Morze	20	M	NS	274	Negative
12	MC00011	0578045	Phyllis	Richards	78	F	NS	303	Positive
13		3319281	Piper	Shaffer	20	F	NS	420	Negative
14		1007104	JoAnn	Simmons	23	F	NS	195	Positive
15		3302562	Amy	Campos	66	F	NS	205	Negative
16		4464119	Armanita	Smith	26	F	NS	223	Positive
17		4465187	Hedda	Chan	46	M	NS	177	Positive
18		0790791	Claire	Yang	45	F	NS	205	Positive
19		0704388	Jamila	Eaton	51	F	NS	430	Positive
20		4284473	Emma	Tom	72	F	NS	182	Negative
21		4253958	Joann	Root	51	F	NS	231	Positive

	A	B	C	D	E	F	G	H	I
1	Master Code	Medical Record	First Name	Last Name	Age	Gender	Tobacco	Total Cholesterol	Fam Hx
2	MC00001	0793073	George	Thompson	35	M	NS	211	Positive
3	MC00002	0657479	Hary	Hagen	66	M	NS	128	Positive
4	MC00003	4400769	Molly	Walton	37	F	NS	163	Positive
5	MC00004	0501127	Marie	Cobb	44	F	NS	142	Negative
6	MC00005	4370320	Bob	Day	37	M	FS	227	Positive
7	MC00006	3154883	Sally	Hahn	42	F	NS	302	Negative
8	MC00007	4522135	Christy	Kelly	33	F	NS	251	Positive
9	MC00008	3044140	Sharon	Lowe	30	F	NS	261	Positive
10	MC00009	0758076	Terri	Mitchell	37	F	NS	310	Negative
11	MC00010	0704755	Toni	Morze	20	M	NS	274	Negative
12	MC00011	0578045	Phyllis	Richards	78	F	NS	303	Positive
13		3319281	Piper	Shaffer	20	F	NS	420	Negative
14		1007104	JoAnn	Simmons	23	F	NS	195	Positive
15		3302562	Amy	Campos	66	F	NS	205	Negative
16		4464119	Armanita	Smith	26	F	NS	223	Positive
17		4465187	Hedda	Chan	46	M	NS	177	Positive
18		0790791	Claire	Yang	45	F	NS	205	Positive
19		0704388	Jamila	Eaton	51	F	NS	430	Positive
20		4284473	Emma	Tom	72	F	NS	182	Negative
21		4253958	Joann	Root	51	F	NS	231	Positive

	A	B	C	D	E	F	G	H	I
1	Master Code	Medical Record	First Name	Last Name	Age	Gender	Tobacco	Total Cholesterol	Fam Hx
2	MC00001	0793073	George	Thompson	35	M	NS	211	Positive
3	MC00002	0657479	Hary	Hagen	66	M	NS	128	Positive
4	MC00003	4400769	Molly	Walton	37	F	NS	163	Positive
5	MC00004	0501127	Marie	Cobb	44	F	NS	142	Negative
6	MC00005	4370320	Bob	Day	37	M	FS	227	Positive
7	MC00006	3154883	Sally	Hahn	42	F	NS	302	Negative
8	MC00007	4522135	Christy	Kelly	33	F	NS	251	Positive
9	MC00008	3044140	Sharon	Lowe	30	F	NS	261	Positive
10	MC00009	0758076	Terri	Mitchell	37	F	NS	310	Negative
11	MC00010	0704755	Toni	Morze	20	M	NS	274	Negative
12	MC00011	0578045	Phyllis	Richards	78	F	NS	303	Positive
13	MC00012	3319281	Piper	Shaffer	20	F	NS	420	Negative
14	MC00013	1007104	JoAnn	Simmons	23	F	NS	195	Positive
15	MC00014	3302562	Amy	Campos	66	F	NS	205	Negative
16	MC00015	4464119	Armanita	Smith	26	F	NS	223	Positive
17	MC00016	4465187	Hedda	Chan	46	M	NS	177	Positive
18	MC00017	0790791	Claire	Yang	45	F	NS	205	Positive
19		0704388	Jamila	Eaton	51	F	NS	430	Positive
20		4284473	Emma	Tom	72	F	NS	182	Negative
21		4253958	Joann	Root	51	F	NS	231	Positive

- If there was more than one level of randomization in your study (anything above a simple random sample) consider using the “concatenate” function in excel to save time creating your unique IDs if they will be a combination of two or more existing columns (e.g. Observation number combined with treatment ID). For example, if you are conducting a 2x2 factorial experiment you could use the following master coding approach:

Formula Bar: `=CONCATENATE(D2,E2,C2)`

	A	B	C	D	E	F	G	H	I
1	Master Code	Medical Record		Factor A	Factor B	First Name	Last Name	Age	Gender
2	HS02	0793073	01	H	S	George	Thompson	35	M
3	HS02	0657479	02	H	S	Hary	Hagen	66	M
4	HM03	4400769	03	H	M	Molly	Walton	37	F
5	HM04	0501127	04	H	M	Marie	Cobb	44	F
6	LS05	4370320	05	L	S	Bob	Day	37	M
7	LS06	3154883	06	L	S	Sally	Hahn	42	F
8	LM07	4522135	07	L	M	Christy	Kelly	33	F
9	LM08	3044140	08	L	M	Sharon	Lowe	30	F
10	NS09	0758076	09	N	S	Terri	Mitchell	37	F
11	NS10	0704755	10	N	S	Toni	Morze	20	M
12	NM11	0578045	11	N	M	Phyllis	Richards	78	F
13	NM12	3319281	12	N	M	Piper	Shaffer	20	F

**Note: If you use this approach be sure that you have enough observations to ensure that non-PHI variables (such as gender and age) cannot be combined to identify individuals: in the above example there are many more individuals in each category (i.e. HS, HM, LS, LM, NS, NM have many more than two individuals each).*

- Step 2: Identify the PHI in your data set
 - In your excel spreadsheet, color the columns that contain PHI in yellow to make them easily visible (see pic below).

	A	B	C	D	E	F	G	H	I
1	Master Code	Medical Record	Name	Age	Gender	Tobacco	Total Cholesterol	Fam Hx	
2	MC00001	0793073	George Thompson	35	M	NS	211	Positive	
3	MC00002	0657479	Hary Hagen	66	M	NS	128	Positive	
4	MC00003	4400769	Molly Walton	37	F	NS	163	Positive	
5	MC00004	0501127	Marie Cobb	44	F	NS	142	Negative	
6	MC00005	4370320	Bob Day	37	M	FS	227	Positive	
7	MC00006	3154883	Sally Hahn	42	F	NS	302	Negative	
8	MC00007	4522135	Christy Kelly	33	F	NS	251	Positive	
9	MC00008	3044140	Sharon Lowe	30	F	NS	261	Positive	
10	MC00009	0758076	Terri Mitchell	37	F	NS	310	Negative	
11	MC00010	0704755	Toni Morze	20	M	NS	274	Negative	
12	MC00011	0578045	Phyllis Richards	78	F	NS	303	Positive	

- The following 18 types of identifiers of the individual or of relatives, employers, or household members of the individual, are PHI:

“(A) Names

- (B) All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census:
 - (1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and
 - (2) The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000
- (C) All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
- (D) Telephone numbers
- (L) Vehicle identifiers and serial numbers, including license plate numbers
- (E) Fax numbers
- (M) Device identifiers and serial numbers
- (F) Email addresses
- (N) Web Universal Resource Locators (URLs)
- (G) Social security numbers
- (O) Internet Protocol (IP) addresses
- (H) Medical record numbers
- (P) Biometric identifiers, including finger and voice prints
- (I) Health plan beneficiary numbers
- (Q) Full-face photographs and any comparable images
- (J) Account numbers
- (R) Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section [Paragraph (c) is presented below in the section “Re-identification”]; and
- (K) Certificate/license numbers”

<https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>

• Step 3: Create Master Code List

- Open a new excel spreadsheet and name it “Master Codes” followed by something to identify the database (e.g. “Master Codes Behavior Study 2019” or “Master Codes Diabetes”, etc.).
- Copy the column from your original database that contains your master codes and insert it into column A in your new spreadsheet (be careful to COPY the column so that it is available in both the new spreadsheet and the old one; see below).

Master Code	First Name	Last Name	Age	Gender	Tobacco	Total Cholesterol	Family
MC0000	George	Thompson	35	M	NS	211	Posit
MC0000	Hary	Hagen	66	M	NS	128	Posit
MC0000	Molly	Walton	37	F	NS	163	Posit
MC0000	Marie	Cobb	44	F	NS	142	Negat
MC0000	Bob	Day	37	M	FS	227	Posit
MC0000	Sally	Hahn	42	F	NS	302	Negat
MC0000	Christy	Kelly	33	F	NS	251	Posit
MC0000	Sharon	Lowe	30	F	NS	261	Posit
MC0000	Terri	Mitchell	37	F	NS	310	Negat
MC0001	Toni	Morze	20	M	NS	274	Negat
MC0001	Phyllis	Richards	78	F	NS	303	Posit
MC0001	Piper	Shaffer	20	F	NS	420	Negat
MC0001	JoAnn	Simmons	23	F	NS	195	Posit
MC00014	Amy	Campos	66	F	NS	205	Negat
MC00015	Armanita	Smith	26	F	NS	223	Posit
MC00016	Hedda	Chan	46	M	NS	177	Posit

Book1 - Excel

Master Code	Medical Record	First Name	Last Name	Age	Gender	Tobacco	Total Cholesterol	Fam
MC00001	0793073	George	Thompson	35	M	NS	211	Posit
MC00002	0657479	Hary	Hagen	66	M	NS	128	Posit
MC00003	4400769	Molly	Walton	37	F	NS	163	Posit
MC00004	0501127	Marie	Cobb	44	F	NS	142	Negat
MC00005	4370320	Bob	Day	37	M	FS	227	Posit
MC00006	3154883	Sally	Hahn	42	F	NS	302	Negat
MC00007	4522135	Christy	Kelly	33	F	NS	251	Posit
MC00008	3044140	Sharon	Lowe	30	F	NS	261	Posit
MC00009	0758076	Terri	Mitchell	37	F	NS	310	Negat
MC00010	0704755	Toni	Morze	20	M	NS	274	Negat
MC00011	0578045	Phyllis	Richards	78	F	NS	303	Posit
MC00012	3319281	Piper	Shaffer	20	F	NS	420	Negat
MC00013	1007104	JoAnn	Simmons	23	F	NS	195	Posit

Master Codes Example - Excel

Master Code							
1	MC00001						
2	MC00002						
3	MC00003						
4	MC00004						
5	MC00005						
6	MC00006						
7	MC00007						
8	MC00008						
9	MC00009						
10	MC00010						
11	MC00011						
12	MC00012						
13	MC00013						
14	MC00014						

- Cut any other columns in yellow out from the original spreadsheet and paste them into the new master code list database (see below).

Book1 - Excel

FileHomeInsertPage LayoutFormulasDataReviewViewAcrobatTell me...Sign inShare

ClipboardFontAlignmentNumberStylesCellsEditing

Calibri11A⁺A⁻

B**I**U

</

Master Code	Age	Gender	Tobacco	Total Cholesterol	Fam
MC00001	35	M	NS	211	Posit
MC00002	66	M	NS	128	Posit
MC00003	37	F	NS	163	Posit
MC00004	44	F	NS	142	Negat
MC00005	37	M	FS	227	Posit
MC00006	42	F	NS	302	Negat
MC00007	33	F	NS	251	Posit
MC00008	30	F	NS	261	Posit
MC00009	37	F	NS	310	Negat
MC00010	20	M	NS	274	Negat

Master Code	Medical Record	First Name	Last Name
MC00001	0793073	George	Thompson
MC00002	0657479	Hary	Hagen
MC00003	4400769	Molly	Walton
MC00004	0501127	Marie	Cobb
MC00005	4370320	Bob	Day
MC00006	3154883	Sally	Hahn
MC00007	4522135	Christy	Kelly
MC00008	3044140	Sharon	Lowe
MC00009	0758076	Terri	Mitchell
MC00010	0704755	Toni	Morze

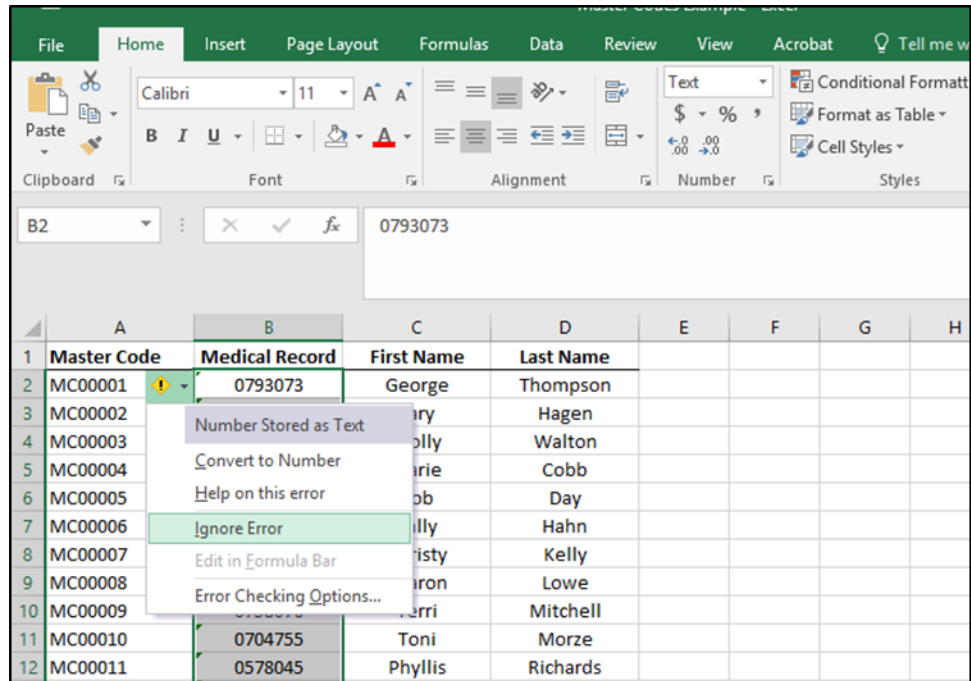
- Delete the columns that are now empty in the original database (in this example columns B,C, and D) and return both databases to having no fill colors using the same method you used to turn columns yellow).
- When you're done with this step you should have both (1) the original data base with the new master codes but no PHI (left) and (2) the new Master Codes database/excel spreadsheet that also contains the master codes and all of the variables that are PHI (right). If done correctly your original database will no longer meet the definition of PHI and thus won't fall under HIPAA jurisdiction and can be shared safely; database 2/Master Codes database is PHI and should be stored and overseen as per HIPAA regulations.

Master Code	Age	Gender	Tobacco	Total Cholesterol	Fam Hx	HTN
MC00001	35	M	NS	211	Positive	N
MC00002	66	M	NS	128	Positive	N
MC00003	37	F	NS	163	Positive	N
MC00004	44	F	NS	142	Negative	Y
MC00005	37	M	FS	227	Positive	N
MC00006	42	F	NS	302	Negative	N
MC00007	33	F	NS	251	Positive	Y
MC00008	30	F	NS	261	Positive	Y
MC00009	37	F	NS	310	Negative	Y
MC00010	20	M	NS	274	Negative	N
MC00011	78	F	NS	303	Positive	Y
MC00012	20	F	NS	420	Negative	N
MC00013	23	F	NS	195	Positive	N
MC00014	66	F	NS	205	Negative	N
MC00015	26	F	NS	223	Positive	Y
MC00016	46	M	NS	177	Positive	N

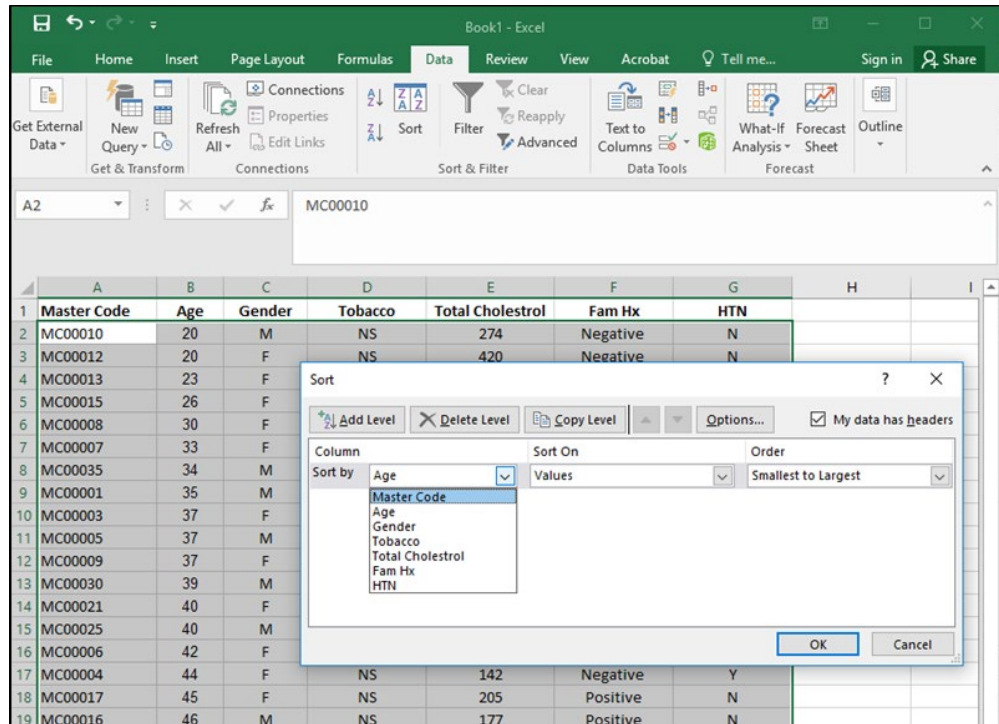
Master Code	Medical Record	First Name	Last Name
MC00001	0793073	George	Thompson
MC00002	0657479	Hary	Hagen
MC00003	4400769	Molly	Walton
MC00004	0501127	Marie	Cobb
MC00005	4370320	Bob	Day
MC00006	3154883	Sally	Hahn
MC00007	4522135	Christy	Kelly
MC00008	3044140	Sharon	Lowe
MC00009	0758076	Terri	Mitchell
MC00010	0704755	Toni	Morze
MC00011	0578045	Phyllis	Richards
MC00012	3319281	Piper	Shaffer
MC00013	1007104	JoAnn	Simmons
MC00014	3302562	Amy	Campos
MC00015	4464119	Armanita	Smith
MC00016	4465187	Hedda	Chan

- Even if your **database 2** is only two columns, the master codes and the patient/subject names, the database **is PHI and under HIPAA jurisdiction** and as such, as “the covered entity” you must not “use or disclose the code or other means of record identification for any other purpose, and ... not disclose the mechanism for re-identification”.

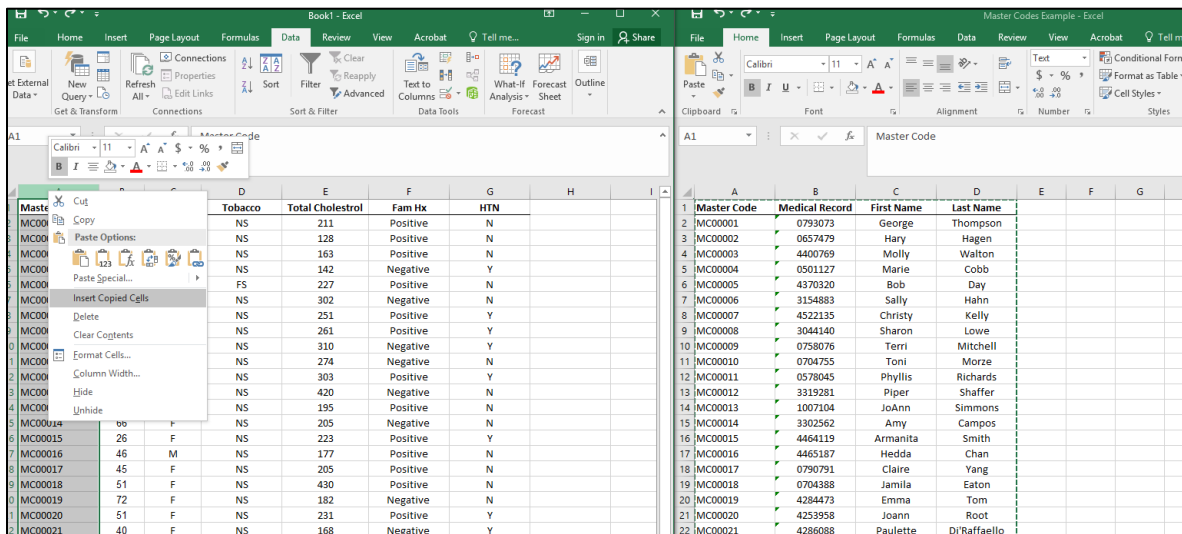
**Note: When numbers are stored as text in excel an error message can occur (e.g. see green triangles on left side of the Medical Record column cells). To get rid of this error message, highlight everything in the column that has the error message, hover the curser over the green arrow, and choose “ignore error” (see below).*



- Step 4: Re-identify the PHI in your data set
 - If you need to re-identify your original database this can be done via the following method:
 - Sort both databases by the unique identification code given by the Master Code.
 - To do this, go to the “Data” tab, choose “Sort”, click “my data has headers”, and then choose “Master Code” (or your column name). Again, be sure to do this for both databases.



- As long as no rows (patients/subjects) have been added or deleted since you created the Master Code database you can copy the PHI data in database 2 and insert it back into database 1 without problems.



- To double-check that the merge was successful, insert a column after the “Master Code” column, and use the following code, =IF(cell address 1 with master code=cell address 2 with master code, “Y”, “9999999”) but be sure to change the column names to match the columns in your database (the two examples below show how the code changes for different column numbers):

File Home Insert Page Layout Formulas Data Review View Acrobat Tell me... Sign in Share									
Get External Data New Query Refresh All Connections Properties Edit Links Sort Filter Clear Reapply Advanced Text to Columns What-If Analysis Forecast Outline									
Get & Transform Connections Sort & Filter Data Tools Forecast									
F2 X ✓ fx =IF(E2=A2,"Y","9999999")									
	A	B	C	D	E	Formula Bar	G	H	I
1	Master Code	Medical Record	First Name	Last Name	Master Code	Merge Correct?	Age	Gender	Toba
2	MC00001	0793073	George	Thompson	MC00001	Y	35	M	N
3	MC00002	0657479	Hary	Hagen	MC00002	Y	66	M	N
4	MC00003	4400769	Molly	Walton	MC00003	Y	37	F	N
5	MC00004	0501127	Marie	Cobb	MC00004	Y	44	F	N
6	MC00005	4370320	Bob	Day	MC00005	Y	37	M	FS
7	MC00006	3154883	Sally	Hahn	MC00006	Y	42	F	N
8	MC00007	4522135	Christy	Kelly	MC00007	Y	33	F	N
9	MC00008	3044140	Sharon	Lowe	MC00008	Y	30	F	N
10	MC00009	0758076	Terri	Mitchell	MC00009	Y	37	F	N
11	MC00010	0704755	Toni	Morze	MC00010	Y	20	M	N
12	MC00011	0578045	Phyllis	Richards	MC00011	Y	78	F	N
13	MC00012	3319281	Piper	Shaffer	MC00013	9999999	20	F	N
14	MC00013	1007104	JoAnn	Simmons	MC00014	9999999	23	F	N
15	MC00014	3302562	Amy	Campos	MC00015	9999999	66	F	N
16	MC00015	4464119	Armanita	Smith	MC00016	9999999	26	F	N

File Home Insert Page Layout Formulas Data Review View Acrobat Tell me... Sign in Share									
Get External Data New Query Refresh All Connections Properties Edit Links Sort Filter Clear Reapply Advanced Text to Columns What-If Analysis Forecast Outline									
Get & Transform Connections Sort & Filter Data Tools Forecast									
QRT X ✓ fx =IF(F2=A2,"Y","9999999")									
	A	B	C	D	E	F	G	H	I
1	Master Code	Medical Record	First Name	Last Name	Age	Master Code	Merge Correct?	Age	Gender
2	MC00001	0793073	George	Thompson	35	MC00001	"", "9999999")	35	M
3	MC00002	0657479	Hary	Hagen	66	MC00002	Y	66	M
4	MC00003	4400769	Molly	Walton	37	MC00003	Y	37	F
5	MC00004	0501127	Marie	Cobb	44	MC00004	Y	44	F
6	MC00005	4370320	Bob	Day	37	MC00005	Y	37	M
7	MC00006	3154883	Sally	Hahn	42	MC00006	Y	42	F
8	MC00007	4522135	Christy	Kelly	33	MC00007	Y	33	F
9	MC00008	3044140	Sharon	Lowe	30	MC00008	Y	30	F
10	MC00009	0758076	Terri	Mitchell	37	MC00009	Y	37	F
11	MC00010	0704755	Toni	Morze	20	MC00010	Y	20	M
12	MC00011	0578045	Phyllis	Richards	78	MC00011	Y	78	F
13	MC00012	3319281	Piper	Shaffer	20	MC00013	9999999	20	F
14	MC00013	1007104	JoAnn	Simmons	23	MC00014	9999999	23	F
15	MC00014	3302562	Amy	Campos	66	MC00015	9999999	66	F
16	MC00015	4464119	Armanita	Smith	26	MC00016	9999999	26	F
17	MC00016	4465187	Hedda	Chan	46	MC00017	9999999	46	M
18	MC00017	0790791	Claire	Yang	45	MC00018	9999999	45	F

- If any “9999999” appear, examine the difference between the two records on that row to see what observation is missing or possibly present in duplicate, and make the most appropriate change in the associated database. If many records are different consider soliciting help from the medical research office to merge the database correctly using statistical software.