**Creating Project Data Sheets**

1. Generally - you don’t ‘pay by the column’ use them – one column, one thought, one type of data
   1. Don’t put multiple bits of information in one column
      1. e.g., indicating multiple diagnoses, complications, etc. use separate columns
      2. Separate fields for any indicator, date, etc.
   2. Yes/no variables should be only that – no reason, no dates, no elaboration 0 and 1 are good for this as codes, if possible
   3. Don’t mix formats in one column– numeric only, dates only, text only
   4. One exception to more columns is better rule: for a mutually exclusive categorical variable – one column with the value clearly indicated, don’t make individual columns with yeses marked under one column – unless it is a check all that apply field
2. When there is missing data
   1. Can always leave blank
   2. If want to indicate to yourself or analyst that the missing info is not an oversight but a known unknown – if numeric, use a number that is impossible to be actual value – 9 maybe or 99 or 9999 etc.; if text, ‘unknown’ is fine; don’t put unknown in a date or numeric field
3. Be extra careful with date columns
   1. Date fields should include only dates – no text, no comments
   2. No impossible dates – like 11/35/2010.
   3. If missing data - code like 99 doesn’t work - DO not use 06/99/2009 if you know it is 06/2009 but don’t know day – if this occurs frequently consider two numeric columns for month and year; if exception, consider using the 15th as an approximation – assuming you are not counting days – like for LOS where it would make a big difference
   4. Date of service is PHI – treat it same as medrec or name – IRB prefers if you treat everything as a timespan (admit is day 0, figure everything else from there – but sometimes not practical or want to use it for day of week/time during the year, etc.
4. Text fields – be careful
   1. Generally, coded fields are better – but understand text clearer to look over
   2. Text okay, best to keep it short
   3. BE CONSISTENT- if you want it to be the same category –use the exact same text
   4. Capitalization and spaces count as being different
   5. Excel will offer text used before – take advantage
5. More generally – Excel has data validation – you can give ranges, create lists, etc. – take advantage
6. Column headings should be one row – each heading unique; SPSS can’t handle multiple lines
7. If you did summary data in Excel – put it in another sheet – don’t use rows for aggregated info
8. Don’t use color to signify anything - SPSS doesn’t read color
9. No need to separate groups into separate worksheets – have an indicator column for group. If you need separate for Excel purposes – keep the indicator variable
10. If more than one person is filling in data – talk before and check back with each other
11. TALK WITH SENIOR SCIENTIST or whomever will be analyzing before you start; then complete about 10 rows, share and talk again

And of course, be HIPAA compliant

1 . Once there is PHI – use HH network computers

2. Avoid laptops, flash drives

3. E-mail okay within network or to UCHC protected addresses – DON”T USE G-MAIL