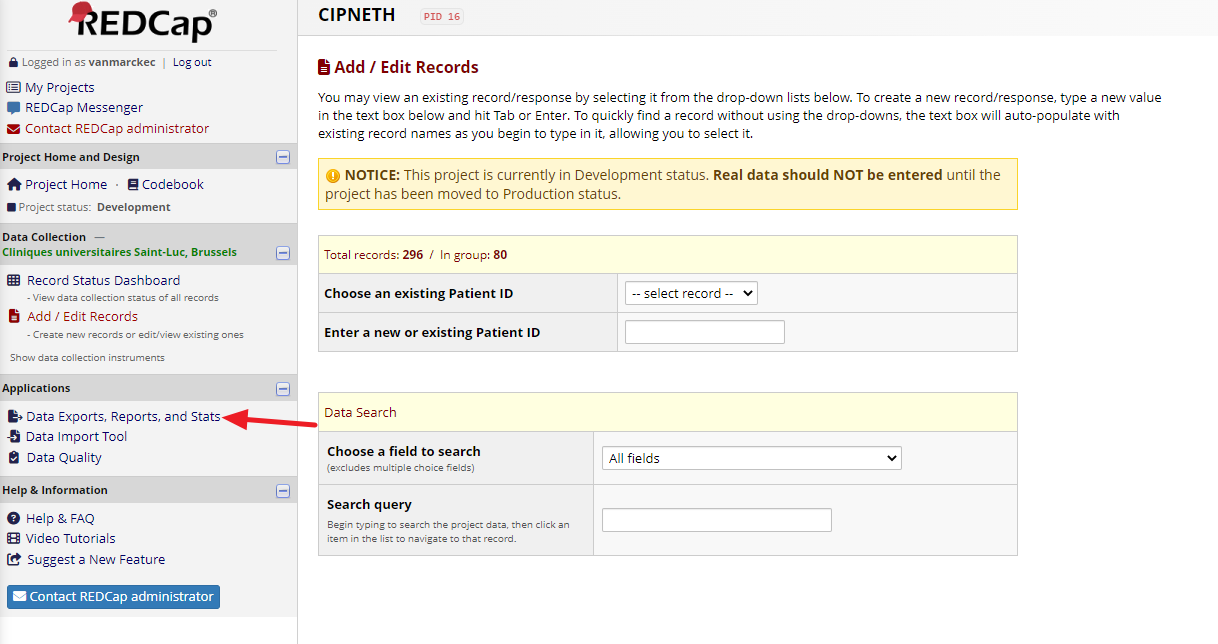
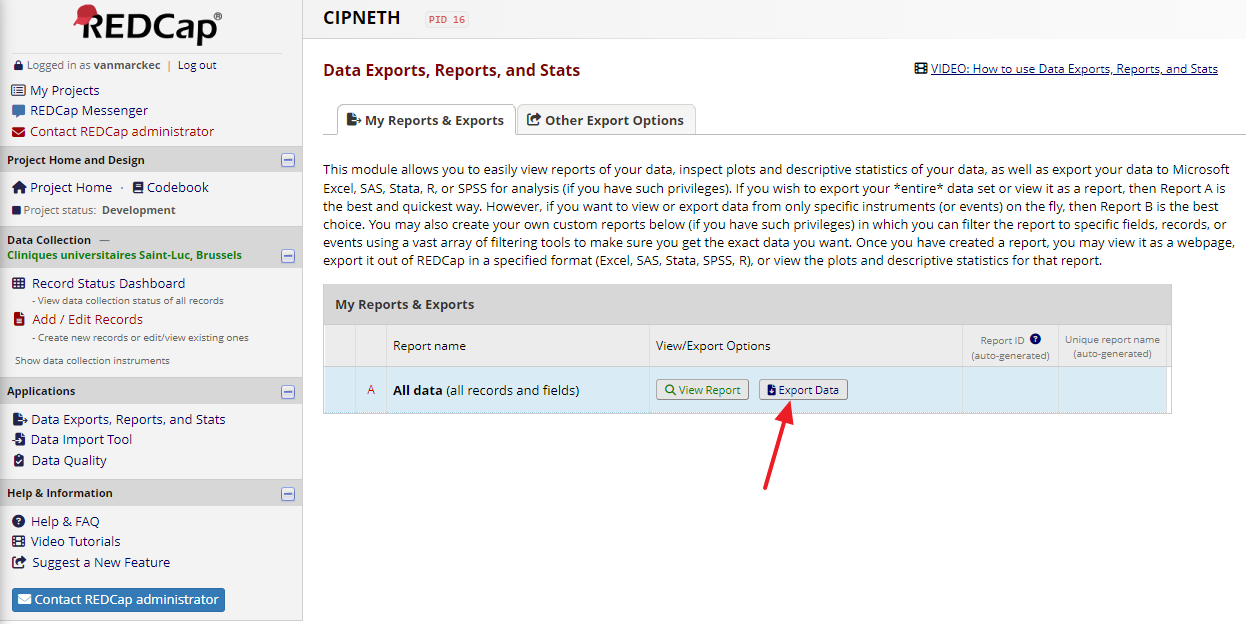
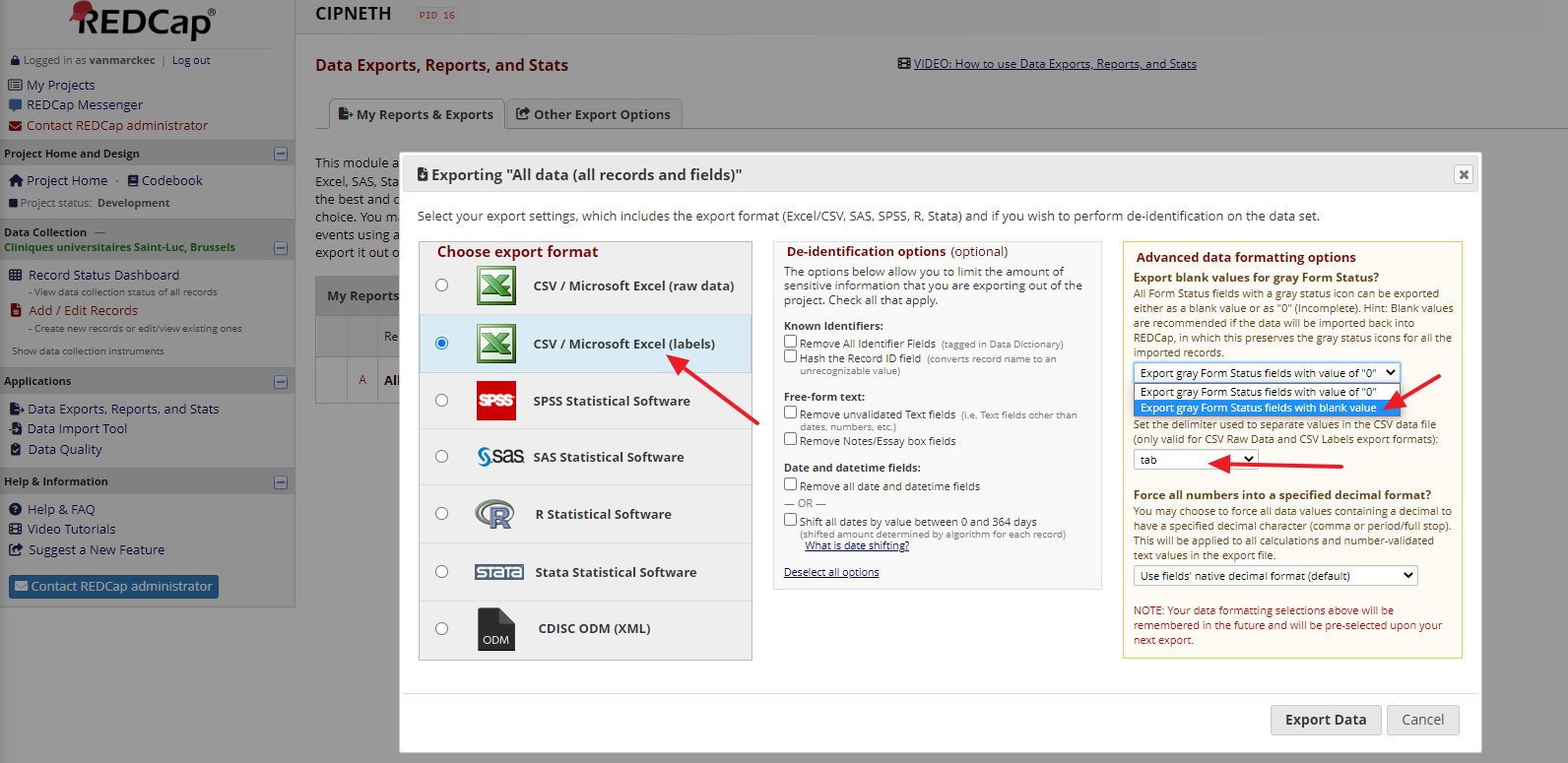
1. Export your data if filled in Redcap, or use your raw file







1. Open the CSV file in excel, R, …
2. Filter on Patient ID (column A) to keep only the “non blanks” (so you see only the unique patients as rows, and not the information regarding each paclitaxel cycle).
3. Hereunder a proposition of quality check to make. Personally, I note each excel cell for which I see a potential issue or a missing data, before going at the end back to the EMR and RedCap. In excel, easiest is to use data>filter of each column. In R, easiest is probably to perform table(dataframe$variable1, dataframe$variable2)
4. Do not forget to make each correction on RedCap and save each screen !

* For each column, check by filtering if there are missing data
* Column G (date of diagnosis) : years in the expected range ?
* Column H or I (age at diagnosis) : in the expected range ?
* Column J (menopausal status) : filter on “pre-menopausal” : look at column H or I (age at diagnosis) : in the expected range ?
* Column L (BSA) : in the expected range ?
* Column AA (clinical T stage) : filter on each possible answer and check if corresponding tumor sizes (column AB) match.
* Column AD (inflammatory breast cancer) : filter on the 2 possible answers and check if Column AA (clinical T stage) matches (T4d ?)
* Column AK (biological subtype) : filter on each possible answer and check if the prognostic factors (columns AF-AG-AH for ER-PR-HER2) match
* Columns BY or CA, and BZ or CB (time from diagnosis to AC and duration of AC) in expected range ?
* Column CC (number of AC cycles) in expected range ?
* Column CD, CE and CF (total cumulative doses of doxo, epi, cycloph) in expected ranges ?
* Columns CH and CI (trastu – pertu neoadj) :
  + Filter on CH “received” : all CI “not received” ? and all AK (bio subtype) are HER2 pos ?
  + Filter on CI “received” : all CH “not received” ? and all AK (bio subtype) are HER2 pos ?
  + Filter on AK (biological subtype) “non luminal B HER2 positive” and “luminal B HER2 positive” : columns CH and CI have the expected values ?
* Column CG (neoadj carboplatin) : filter on “received” : all AK (bio subtype) are triple negative ?
* Columns CK and CL (neuropathy and date of occurrence) : if “presence”, marked date ? If marked date, “presence” and not “absence” ?
* Column CM or CN (time to neuropathy) : in expected ranges ?
* Column CP (pre existing neuropathy) : if “presence”, CK (neuropathy during pacli) marked as “presence” ?
* Column CV or CW (time to surgery) : in expected ranges ?
* Column CX and CY (types of surgery) : mutually exclusive ? (NB : could in rare cases not be exclusive, if re-excision).
* Columns CZ, DA, DB (types of ax dissection) : mutually exclusive ? (NB : could in rare cases not be exclusive).
* Columns DC (pathological T stage) : filter on each possible answer and check if corresponding tumor sizes (column DD) match.
* Column DF (pCR) : filter on the 2 possible values and check if columns DC and DE match.
* Column DI (adj capecitabine) : filter on “received” and check if AK (bio subtype) is TNBC
* Column DJ (adj emtansine) : filter on “received” and check if AK (bio subtype) is HER2 pos
* Column DJ (adj emtansine) : filter on “received” and check if :
  + DF (pCR) is “No”
  + AK (bio subtype) is HER2 positive (rare cases of HER2 negative at diagnosis that should then be HER2 positive after neoadj treatment by clonal selection)
* Column DK (adj trastu without pertu) : filter on “received” and check if :
  + DL (adj trastu with pertu) is “not received”
  + AK (bio subtype) is HER2 positive (rare cases of HER2 negative at diagnosis that should then be HER2 positive after neoadj treatment by clonal selection)
* Column DL (adj trastu with pertu) : filter on “received” and check if :
  + DK (adj trastu without pertu) is “not received”
  + AK (bio subtype) is HER2 positive (rare cases of HER2 negative at diagnosis that should then be HER2 positive after neoadj treatment by clonal selection)
* Column DN (endocrine therapy) : filter on “received” and check if AK (bio subtype) is “luminal B HER2 positive” (rare cases of HR negative at diagnosis that should then be HR positive after neoadj treatment by clonal selection)
* Columns DR or DS (time between diagnosis and vital status) : in expected ranges ?
* Column DT (date of the first IBC-free event). Filter on “not blank” : at least one of the columns DW or DZ (date of local recurrence, date of distant recurrence) should be filled.