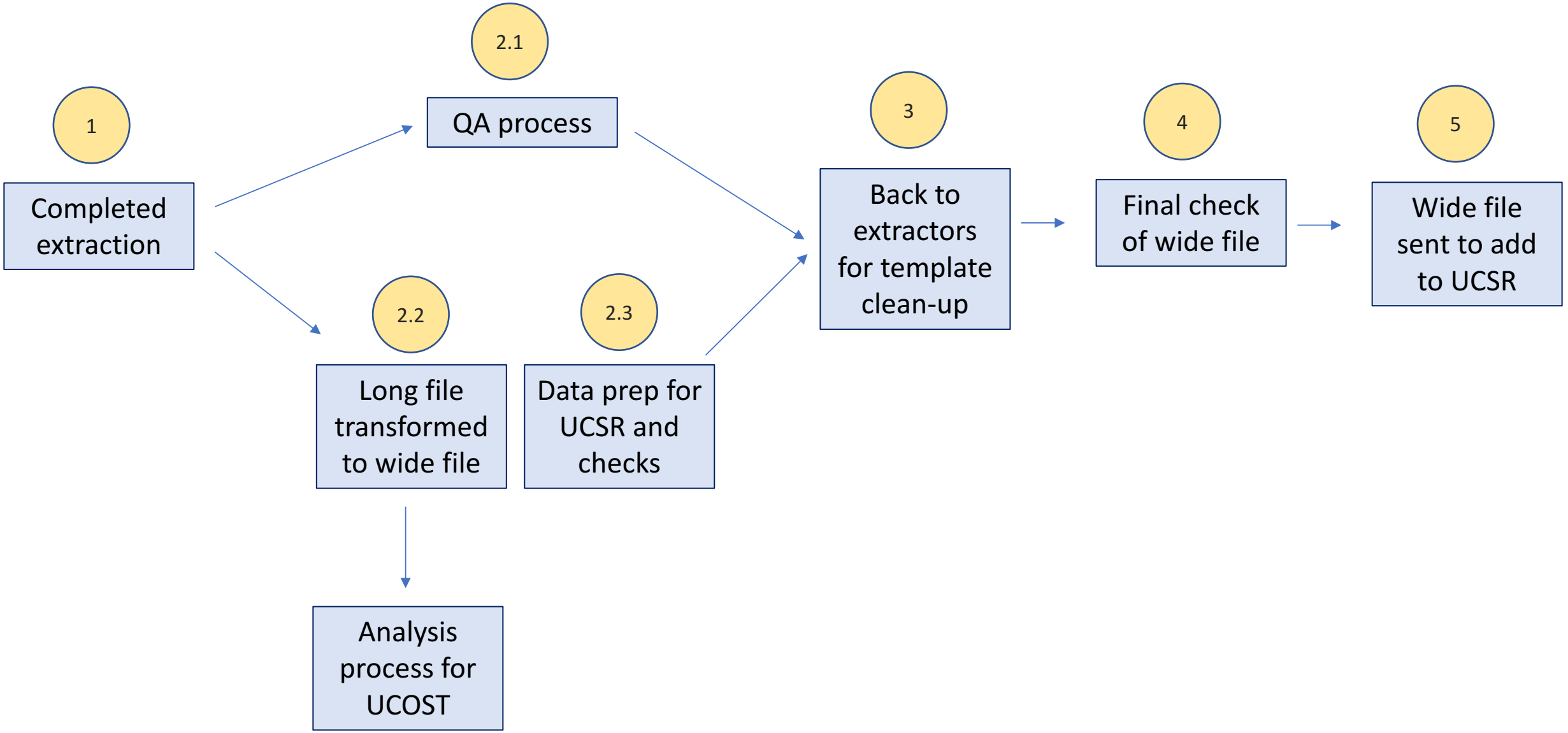
Extraction → UCSR

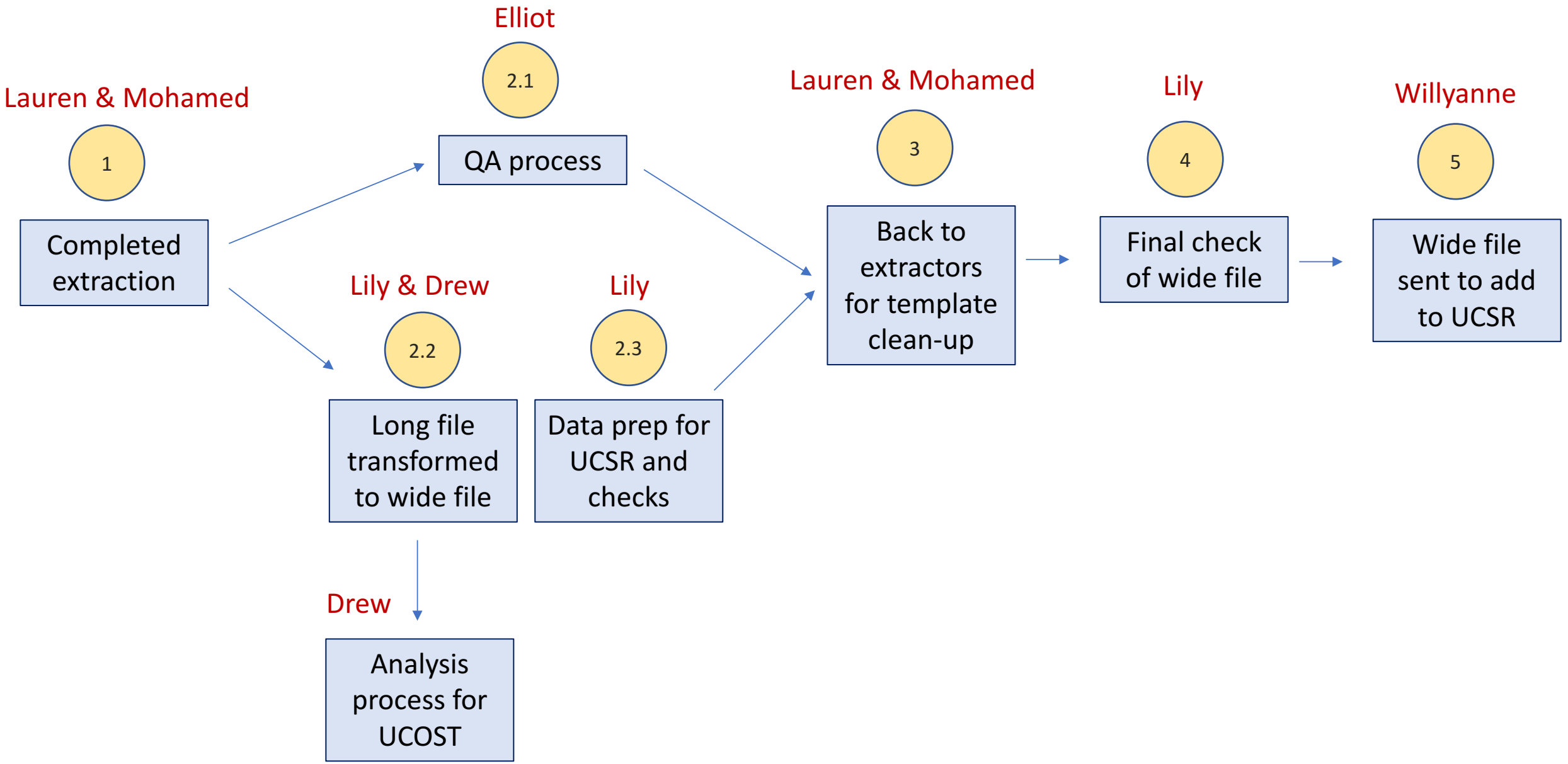
February 26, 2018

Overview

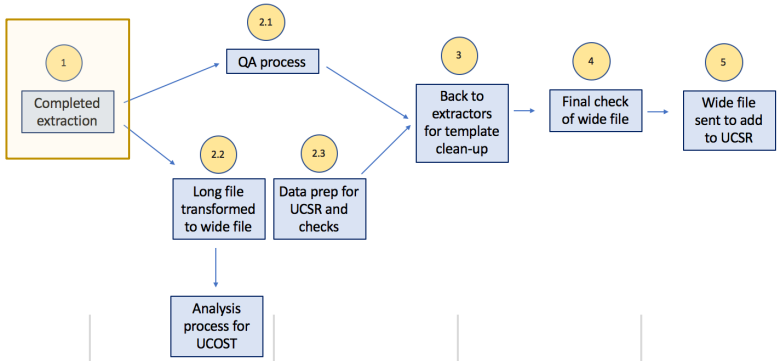
Activity	Responsible team	Average time to complete
Data Extraction	Lauren & Mohamed	14 articles/wk. Expect May 4 finish
Quality Assurance	Elliot	1.25 hours/article
Long file → wide file*	Lily & Drew	~2 days/intervention
Data prep for UCSR and checks*	Lily	1 day/intervention
Data Clean Up	Lauren (& Mohamed)	~1 day/intervention

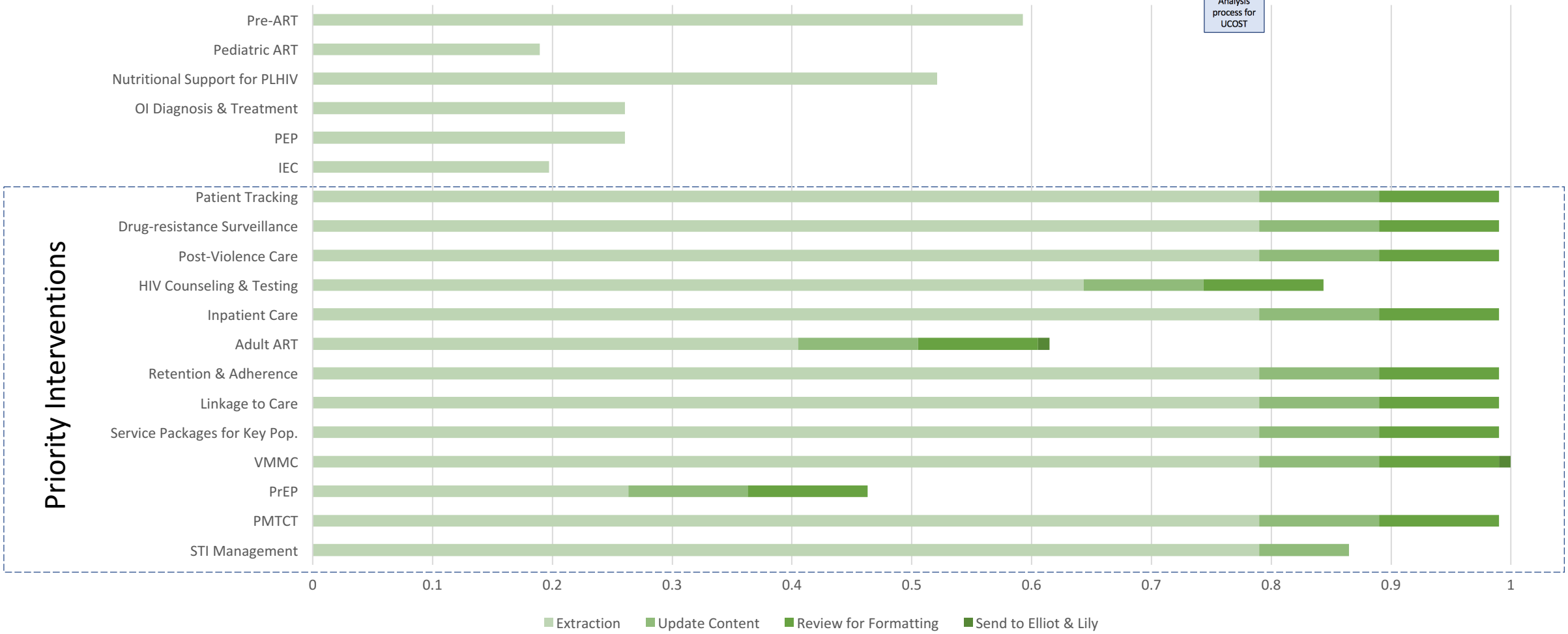
* These processes take the above stated time to complete when they are being adapted for a new intervention. Re-running these processes takes minutes.


1
Completed extraction
2.1
QA process
2.2
Long file transformed to wide file
2.3
Data prep for UCSR and checks
Analysis process for UCOST
3
Back to extractors for template clean-up
4
Final check of wide file
5
Wide file sent to add to UCSR

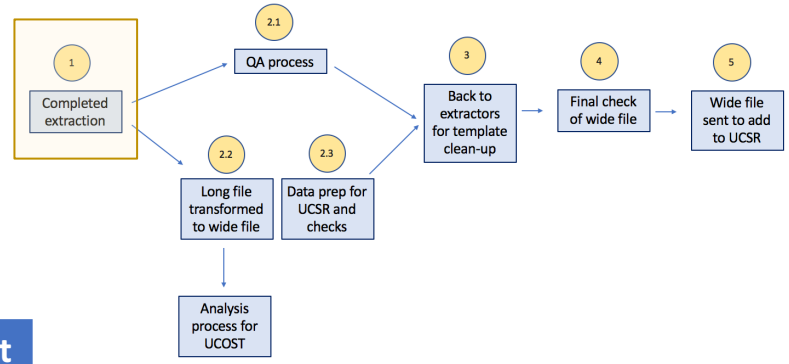

Lauren & Mohamed
1
Completed extraction
Elliot
2.1
QA process
Lily & Drew
2.2
Long file transformed to wide file
Drew
Analysis process for UCOST
Lily
2.3
Data prep for UCSR and checks
Lauren & Mohamed
3
Back to extractors for template clean-up
Lily
4
Final check of wide file
Willyanne
5
Wide file sent to add to UCSR

1 Completed extraction





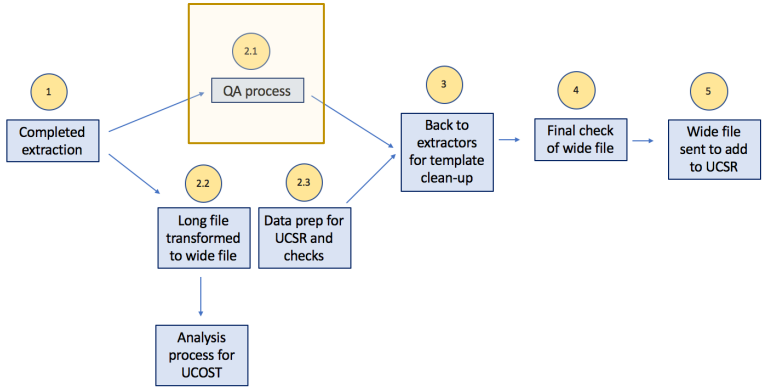
1 Completed extraction



Intervention	Articles & Reports to Extract
Condom Social Marketing	10
Male Condom Provision	1
Female Condom Provision	1
Workplace Safety Package	4
Blood Safety	1
Injection Safety	3
Needle & Syringe Programs	8
Opioid Substitution Therapy	13
OI Prophylaxis	7
Viral Load Monitoring	7
HIV/TB Care Delivery	15
Provider Engagement/Training	1
Supply Chain Management	3

2.1

QA Process

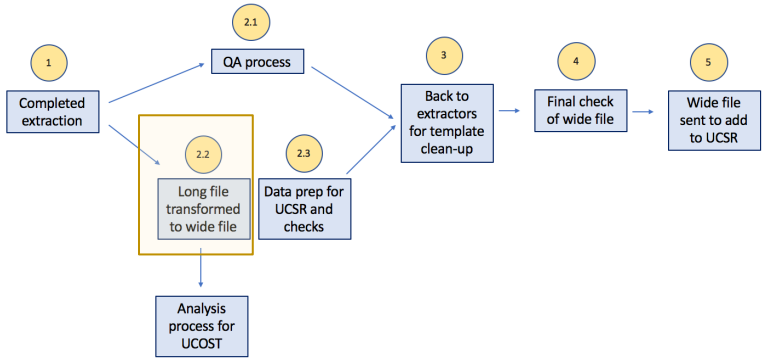


- Elliot reads each article for a given intervention and reviews relevant QA fields (those used in the UCSR)
- Elliot creates new columns next to each QA field for notes
- Elliot uses code for any edits (e.g. to change ‘urban’ to ‘rural’ in “Urbanicity” field Elliot notes ‘@rural’) in the relevant adjacent cell
- As needed, Elliot consults with extraction team for clarifications

Once complete, Elliot sends QA’d dataset to primary extractor for review of proposed edits

2.2

Long file transformed to wide file



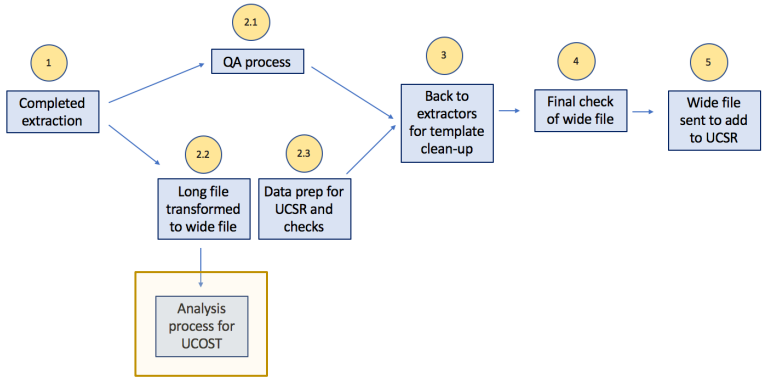
- Code has been written to perform this process. It must be adapted for each intervention. The code includes the following elements:
  - Inflate prices to 2016 dollars using US CPI
  - Cycle through input cost categories (as reported, standard input, activity) and create binaries for reshape process
  - Initial cleaning of study attributes variables and creation of variables to be used in analysis (e.g. meta facility category)
  - Collapse input cost categories and merge with broad cost category so that in wide format and merge with study attributes

Output wide file as a Stata dta for concurrent analysis process and data prep

2.2

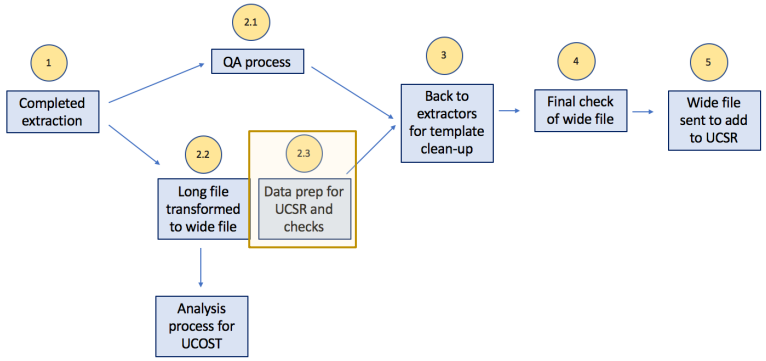
Analysis process for UCOST

- Drew



2.3

Data prep for UCSR & checks

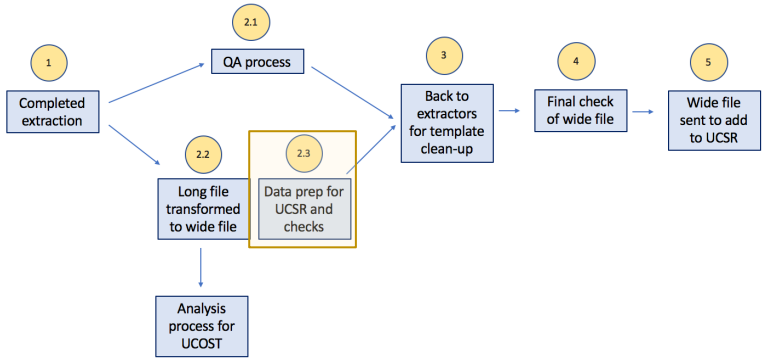


- Cross-validation of costs
  - Check that broad standard input categories sum to mean cost, narrow input categories sum to broad categories, broad activity categories sum to mean cost, narrow input categories sum to broad categories
- Label all variables that will be used in UCSR
- Add in any missing columns for particular intervention that need to be accounted for in UCSR
- Standardize all types of missing values (N/A, NR, NA) with a period
- Order variables for UCSR (inherent check that all variables are there)

Output wide file as .xlsx

2.3

Data prep for UCSR & checks



- Use prepped excel file to generate a pdf of checks
  - Summary of every variable going to the UCSR (categories, labels, frequencies)

```
iso_code                                                    Reported Currency
-------------------------------------------------------------------------------

                type:  string (str6)

        unique values:  4                       missing "":  0/101

           tabulation:  Freq.  Value
                           1  "Kwacha"
                           1  "Maloti"
                          90  "USD"
                           9  "ZAR"
```

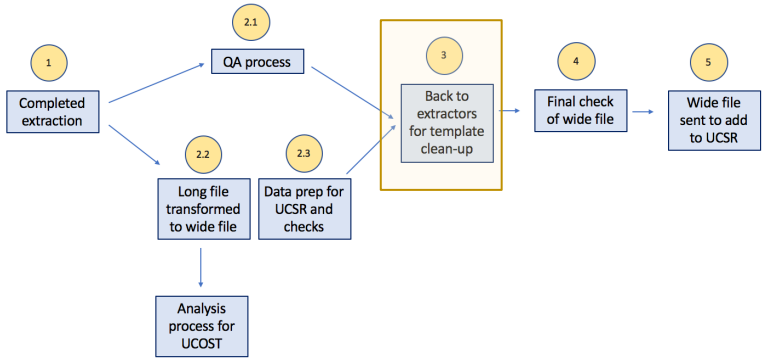
```
-------------------------------------------------------------------------------
id_pop_dem_std                                      Target group (demographic)
-------------------------------------------------------------------------------

                type:  string (str41)

        unique values:  5                       missing "":  0/101

           tabulation:  Freq.  Value
                          11  "Adolescent Boys and Adult Men"
                           2  "Adolescent Boys and Young Men, ages
                              15-29"
                           5  "Infants, ages 1month - 2 yrs"
                          82  "Men"
                           1  "Neonates, ages birth - 1 month"

             warning:  variable has embedded blanks
```

What is the difference between:

"Adolescent boys and adult men"
"Adolescent boys and young men, ages 15-29"
"Men"

3

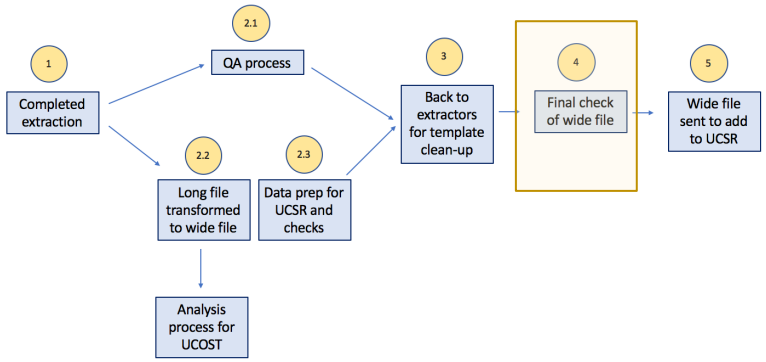
Back to extractors for template clean-up



- Lily sends data checks report with comments
  - Primary extractor(s) review Lily's comment and address any errors
- Elliot sends QA'd data with proposed changes
  - Primary extractor(s) review changes for clarity
  - Data goes to Drew/Lily to implement changes using Stata code
- Lauren spot-checks datasheet for outstanding issues before sending to Lily

4

Final check of wide file

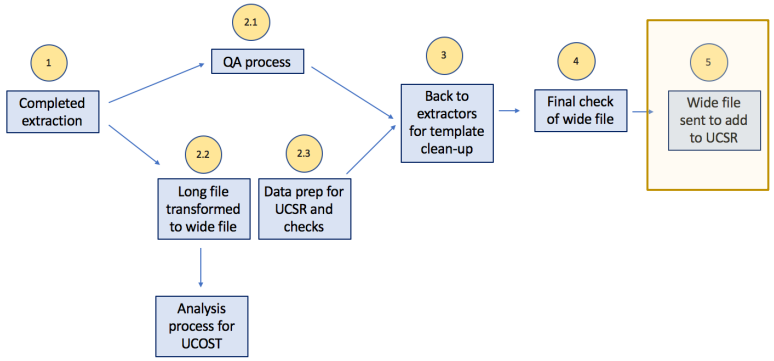


- Lily runs revised extraction sheet back through the UCSR prep process to produce a final wide file

- Conducts last spot-checks for spelling, capitalization, missing values, etc.

5

Wide file sent to add to UCSR



- Lily sends Willyanne wide file after this entire process is completed for a given intervention

- Willyanne reviews the wide file and shares with programmers to upload to UCSR
  - Any questions/comments are sent back to Lily