

FPsim New User Instructions

Parameterizing to a new context

Option A: Using an existing parameter file, e.g. kenya.py

Scalar parameters

Scalar parameters can be changed based on new data. At minimum, change 'location' to your new location.

```
def scalar_pars():
    scalar_pars = {
        'location': 'kenya',
        'postpartum_dur': 23,
        'breastfeeding_dur_mu': 11.4261936291137,
        'breastfeeding_dur_beta': 7.5435309020483,
        'abortion_prob': 0.201,
        'twins_prob': 0.016,
        'high_parity_nonuse': 1,
        'mcpr_norm_year': 2020,
    }
    return scalar_pars
```

1. `Postpartum_dur` → Postpartum duration. This is the number of timesteps (months) that we consider agents to be in the postpartum state. Vary this if you want to consider behavioral change postpartum for longer or shorter timeframes.
2. `Breastfeeding_dur_mu` → Breastfeeding duration mu. This is the mu for the parametric distribution of breastfeeding duration for each live birth. Currently parameterized using the children's recode DHS file, see `data_processing/breastfeeding_stats.R`
3. `Breastfeeding_dur_beta` → Breastfeeding duration beta. This is the beta for the parametric distribution of breastfeeding duration for each live birth. See `Breastfeeding_dur_mu` for more.
4. `abortion_prob` → Per pregnancy probability of abortion. This is currently a single parameter applied to the entire population, based on Guttmacher estimates:
<https://bmcpregnancychildbirth.biomedcentral.com/articles/10.1186/s12884-015-0621-1>
When available and reliable, additional data sources may be used to create a more complex abortion parameter.
5. `twins_prob` → Probability that a live birth will result in twins. From
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0025239>
6. `high_parity_nonuse` → Users can determine whether to adjust the probability of non-use by women with high parity. This parameter is currently used in determining postpartum method choice, in the `people.py` script.

7. mCPR_norm_year → Year of normalization for mCPR trends. Can correspond to the latest year of data.

Defining a method to help build our parameters

Data2Interp → a method we define to convert unevenly spaced data into an even spline interpolation.
No need to change.

```
def data2interp(data, ages, normalize=False):  
    ''' Convert unevenly spaced data into an even spline interpolation '''  
    model = si.interp1d(data[0], data[1])  
    interp = model(ages)  
    if normalize:  
        interp = np.minimum(1, np.maximum(0, interp))  
    return interp
```

Files for calibration

See data processing folder for more on how to create these files.

```
def filenames():
    ''' Data files for use with calibration, etc -- not needed for running a sim '''
    files = {}
    files['base'] = sc.thisdir(aspath=True) / 'kenya'
    files['basic_dhs'] = 'basic_dhs.yaml' # From World Bank https://data.worldbank.org/indicator/SH.STA.MMRT?locations=KE
    files['popsize'] = 'popsize.csv' # Downloaded from World Bank: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=KE
    files['mcpr'] = 'cpr.csv' # From UN Population Division Data Portal, married women 1970-1986, all women 1990-2030
    files['tfr'] = 'kenya_tfr.csv' # From World Bank https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=KE
    files['asfr'] = 'asfr.csv' # From UN World Population Prospects 2022: https://population.un.org/wpp/Download/Standard/Fer
    files['ageparity'] = 'ageparity.csv' # Choose from either DHS 2014 or PMA 2022
    files['spacing'] = 'birth_spacing_dhs.csv'
    files['methods'] = 'mix.csv'
    files['afb'] = 'afb.table.csv'
    files['use'] = 'use.csv'
    return files
```

Base parameters

1. age_pyramid → Population pyramid for initialization. Data from WPP if at national level. Age bins match data source. All can be replaced.

```
# %% Demographics and pregnancy outcome

def age_pyramid():
    """
    Starting age bin, male population, female population
    Data are from World Population Prospects
    https://population.un.org/wpp/Download/Standard/Population/
    """
    pyramid = np.array([[0, 801895, 800503], # Kenya 1960
                        [5, 620524, 625424],
                        [10, 463547, 464020],
                        [15, 333241, 331921],
                        [20, 307544, 309057],
                        [25, 292141, 287621],
                        [30, 247826, 236200],
                        [35, 208416, 190234],
                        [40, 177914, 162057],
                        [45, 156771, 138943],
                        [50, 135912, 123979],
                        [55, 108653, 111939],
                        [60, 85407, 94582],
                        [65, 61664, 71912],
                        [70, 40797, 49512],
                        [75, 22023, 29298],
                        [80, 11025, 17580],
                        ], dtype=float)

    return pyramid
```

2. Urban_proportion → Proportion of the population that is in urban residence. Users need to create an urban.csv using the script at locations/data_processing/empowerment_urban_partnership.py

Note: Users parameterizing to an all-urban or all-rural location can replace this with a single parameter (can place into scalar parameters if urban_proportion=1 or 0 instead of defining the proportion here).

```
def urban_proportion():
    """Load information about the proportion of people who live in an urban setting"""
    urban_data = pd.read_csv(thisdir / 'kenya' / 'urban.csv')
    return urban_data["mean"][0] # Return this value as a float
```

3. Age_mortality → Age-dependent mortality rates

User will need two files – mortality_probs and mortality_trends

Mortality_probs derived from probability of dying each year. Data can be found at:

<https://population.un.org/wpp/>

Mortality trend from crude death rate per 1000 people, also from UN Data Portal, 1950-2030:
<https://population.un.org/dataportal/data/indicators/59/locations/404/start/1950/end/2030/table/pivotbylocation>

Projections go out until 2030, but the csv file can be manually adjusted to remove any projections and stop at your desired year.

Both data files can be created using `locations/data_processing/UN_data_scraping.py`

4. `maternal_mortality` → maternal mortality ratio (MMR) per 100,000 live births, as a time trend

From World Bank indicators for maternal mortality ratio (modeled estimate) per 100,000 live births:
<https://data.worldbank.org/indicator/SH.STA.MMRT>

Users may scrape the data using `locations/data_processing/WorldBank_data_scraping.py`

Note: Maternal mortality ratios are notoriously uncertain – users should consider adding low, medium, and high estimates (see `Senegal.py` for an example of using published estimates)

Optional: odds ratios to increase maternal mortality probability at specific ages, example below for adolescent mothers

```
maternal_mortality['ages'] = np.array([16, 17, 19, 22, 25, 50])
maternal_mortality['age_probs'] = np.array([2.28, 1.63, 1.3, 1.12, 1.0, 1.0])
```

5. `infant_mortality` → Infant mortality per 1000 live births as a time trend

From World Bank indicators for infant mortality (< 1 year) per 1000 live births

Users may scrape the data using `locations/data_processing/WorldBank_data_scraping.py`

Optional: odds ratios to increase adolescent risk of infant mortality. Gradient taken from Noori et al for Sub-Saharan African from 2014-2018. Odds ratios with age 23-25 as reference group:

<https://www.medrxiv.org/content/10.1101/2021.06.10.21258227v1>

6. `miscarriage` → Likelihood of miscarriage, assessed per pregnancy

Returns a linear interpolation of the likelihood of a miscarriage by age, taken from data from Magnus et al BMJ 2019: <https://pubmed.ncbi.nlm.nih.gov/30894356/>

Age 0 and 5 set at 100% likelihood. Age 10 imputed to be symmetrical with probability at age 45 for a parabolic curve

7. `Stillbirth` → Likelihood of stillbirth, assessed at delivery

From Report of the UN Inter-agency Group for Child Mortality Estimation, 2020

<https://childmortality.org/wp-content/uploads/2020/10/UN-IGME-2020-Stillbirth-Report.pdf>

Optional: Odds ratios for adolescents. Age adjustments come from an extension of Noori et al., which were conducted June 2022.

8. `Female_age_fecundity` → underlying biological capacity to conceive, per sexually active month

Data from fecundity rates from PRESTO study:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5712257/>

Fecundity rate assumed to be approximately linear from onset of fecundity around age 10 (average age of menses 12.5) to first data point at age 20. The 45-50 age bin is estimated at 0.10 of fecundity of 25-27 yr olds.

For most contexts, this parameter will not change. Very rare to have context-specific fecundity rates from prospective cohort studies. This parameter SHOULD NOT be estimated from pregnancy outcomes.

9. Fecundity_ratio_nullip → nulliparous adjustment to fecundity

Ratio of fecundity that is applied as women continue to age without conceiving. Fecundity is reduced, as an adjustment for primary infertility. Data is also derived from the PRESTO study. Most users will not change this parameter.

10. Lactational amenorrhea → likelihood of lactational amenorrhea

Returns an array of the percent of breastfeeding women by month postpartum 0-11 months who meet criteria for LAM:

Exclusively breastfeeding (bf + water alone), menses have not returned. Extended out 5-11 months to better match data as those women continue to be postpartum insusceptible.

Based on DHS calendar data. Users can calculate using `locations/data_processing/postpartum_code.R`

11. Sexual activity → Likelihood of being sexually active in given month (0/1)

Returns a linear interpolation of rates of female sexual activity, defined as percentage women who have had sex within the last four weeks. From STAT Compiler DHS <https://www.statcompiler.com/en/>

Using indicator "Timing of sexual intercourse"

Excludes women who answer "never had sex", probabilities are only applied to agents who have sexually debuted

Onset of sexual activity probabilities assumed to be linear from age 10 to first data point at age 15

12. Sexual activity pp → sexual activity postpartum

Returns an array of monthly likelihood of having resumed sexual activity within 0-35 months postpartum (note: NOT age-specific – probabilities are estimated by months since last birth).

Users may use DHS individual recode data, using variables `postpartum (v222)`, months since last birth, and sexual activity within 30 days. Parameter is based on survey-weighted data. Data is weighted.

Note: Postpartum month 0 refers to the first month after delivery

13. Debut age (sexual debut)

Returns an array of weighted probabilities of sexual debut by a certain age 10-45. Data taken from DHS variable `v531` (imputed age of sexual debut, imputed with data from age at first union). Users can refer to `locations/data_processing/sexual_debut_age_probs.py` for integrating data from other DHS countries

14. Exposure age

Assumption-based parameter. Acts as a multiplier to the likelihood of conceiving, by age. Users should keep this parameter at 1 for all ages until they have parameterized the entire model and need to adjust for calibration. Keep this parameter influence to a minimum, if needed at all.

15. Exposure parity

Assumption-based parameter. Acts as a multiplier to the likelihood of conceiving, by parity. Users should keep this parameter at 1 for all ages until they have parameterized the entire model and need to adjust for calibration. Keep this parameter influence to a minimum, if needed at all.

16. Birth spacing preferences

Assumption-based parameter. Acts as a multiplier to the likelihood of sexual activity, by postpartum month. Users should keep this parameter at 1 for all ages until they have parameterized the entire model and need to adjust for calibration. Keep this parameter influence to a minimum, if needed at all.

17. Methods

Parameter to label methods, flag them as modern (True) or traditional (False), and determines efficacy. Efficacy from Polis et al: [Contraceptive Failure Rates in the Developing World: An Analysis of Demographic and Health Survey Data in 43 Countries | Guttmacher Institute](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4970461/) (also available here: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4970461/>)

Contraceptive prevalence normalized by time trends here:

```
cpr_data = pd.read_csv(thisdir / 'kenya' / 'cpr.csv')
methods['mcpr_years'] = cpr_data['year'].to_numpy()
methods['mcpr_rates'] = cpr_data['cpr'].to_numpy() / 100 # convert from percent to rate
```

18. Method_probs

Define "raw" (un-normalized, un-trended) matrices to give transitional probabilities from contraceptive calendar data (DHS or PMA, usually).

Probabilities in this function are annual probabilities of initiating (top row), discontinuing (first column), continuing (diagonal), or switching methods (all other entries).

Probabilities at postpartum month 1 are 1-month transitional probabilities for starting a method after delivery.

Probabilities at postpartum month 6 are 5-month transitional probabilities for starting or changing methods over the first 6 months postpartum.

Calendar data can be integrated using `locations/data_processing/make_matrices_dhs.R` or `make_matrices_pma.R`

19. Barriers to use → barriers to using contraceptives

Barriers drawn from DHS data. Currently un-used in FPsim.

Empowerment Module (Optional)

If you are not using the empowerment module, leave these as-is.

- 20. Empowerment_sexual_autonomy
- 21. Empowerment_decision_wages
- 22. Empowerment_decision_health
- 23. Empowerment_paid_employment
- 24. Empowerment_regression_pars
- 25. Empowerment_distributions
- 26. Age_partnership
- 27. Education_objective
- 28. Education_attainment
- 29. Education_dropout_probs
- 30. Education_distributions