Using `demogsurv` with `rdhs`

Jeff Eaton, Bruno Masquelier, and OJ Watson

2018-06-04

This vignette illustrates use of `demogsurv` and `rdhs` to calculate fertility and mortality indicators for lots of DHS surveys in sub-Saharan Africa, and compare estimates to those produced for the DHS StatCompiler. It is currently a hastily developed analysis script, though may be further developed in the future.

Install and load packages

```
## install.packages("devtools")
## devtools::load_all("OJWatson/rdhs")
## devtools::load_all("mrc-ide/demogsurv")

library(rdhs)
library(demogsurv)
library(ggplot2)
library(data.table)
library(haven)

## a little nugget to return API requests as data.table rather than data.frame.
Sys.setenv(rdhs_DATA_TABLE = "TRUE")
```

Identify surveys and datasets

Identify all DHS surveys conducted in sub-Saharan Africa since the year 2005.

```
countries <- dhs_countries()
cc <- countries[RegionName == "Sub-Saharan Africa"]$DHS_CountryCode
surveys <- dhs_surveys(countryIds = cc, surveyYearStart=2005, surveyType = "DHS")
```

Identify individual recode (IR) and births recode (BR) datasets corresponding to these surveys.

```
ird <- dhs_datasets(fileType = "IR", fileFormat = "flat")[SurveyId %in% surveys$SurveyId]
brd <- dhs_datasets(fileType = "BR", fileFormat = "flat")[SurveyId %in% surveys$SurveyId]
```

Use `rdhs` to retreive datasets, downloading them from DHS website if not already in the `rdhs` cache.

```
ird$path <- unlist(get_datasets(ird$FileName))
brd$path <- unlist(get_datasets(brd$FileName))
```

Load all of the datasets into R as a list.

```
ir <- list()
for(survid in ird$SurveyId){
  print(survid)
  dat <- readRDS(ird[SurveyId == survid]$path)
  dat <- dat[grep("caseid|^v0|^v1|^b|^mm", names(dat))]
  ir[[survid]] <- dat
}
```

```
br <- list()
for(survid in brd$SurveyId){
  print(survid)
  dat <- readRDS(brd[SurveyId == survid]$path)
  dat <- dat[grep("caseid|^v0|^v1|^b", names(dat))]
  br[[survid]] <- dat
}

## Convert to factors (a bit inefficient)
ir <- lapply(ir, haven::as_factor)
br <- lapply(br, haven::as_factor)

## Add survey-level variables
ir <- Map(data.frame,
          SurveyId = surveys$SurveyId,
          CountryName = surveys$CountryName,
          SurveyYear = surveys$SurveyYear,
          ir)

br <- Map(data.frame,
          SurveyId = surveys$SurveyId,
          CountryName = surveys$CountryName,
          SurveyYear = surveys$SurveyYear,
          br)
```

Note that `rdhs` provides better tools to extract variables and pool datasets which haven't been fully embraced here.

Use `demogsurv` to analyse demographic rate indicators

Fertility

Calcualte TFR and 15-19 ASFR for 3 year period preceding survey (default argument `tips=c(0, 3)`).

```
tfr <- lapply(ir, calc_tfr, by=~SurveyId+CountryName+SurveyYear, strata=NULL)
tfr <- do.call(rbind, tfr)

asfr15to19 <- lapply(ir, calc_asfr, by=~SurveyId + CountryName + SurveyYear,
                     agegr = c(15, 20), strata=NULL)
asfr15to19 <- do.call(rbind, asfr15to19)
```

Adult mortality

Identify surveys that include sibling history model via querying the DHS API survery with "Maternal mortality" characteristic.

```
survchar <- dhs_surveyCharacteristics()
survchar[grepl("Maternal", SurveyCharacteristicName)]

##    SurveyCharacteristicID SurveyCharacteristicName
## 1:                      1       Maternal mortality

mm_surv <- dhs_surveys(surveyCharacteristicIds = 1)
has_mm <- surveys$SurveyId %in% mm_surv$SurveyId
```

Reshape IR datasets to one row per sibling episode, create a binary variable indicating sibling death, and calculate $_{35}q_{15}$ estimates by sexx.

```
sib <- lapply(ir[has_mm], reshape_sib_data,
              widevars = c("SurveyId", "CountryName", "SurveyYear", "v005", "v008", "v021"))
sib <- lapply(sib, function(x){x$death <- factor(x$mm2, c("dead", "alive")) == "dead"; x})
q3515 <- lapply(sib, calc_nqx, by=~SurveyId+CountryName+SurveyYear + mm1, strata = NULL,
                agegr=seq(15, 50, 5), tips=c(0, 7), dob="mm4", dod="mm8")
q3515 <- do.call(rbind, q3515)
```

Child mortality

demogsurv does not yet implement the exact child mortality calculation produced in DHS reports and DHS StatCompiler (see Rutstein and Rojas 2006. This is planned for future implementation.

The function calc_nqx() calculates piecewise constant mortality rates within age groups 0, 1-2, 3-4, 5-11, 12-24 months, and 2, 3, and 4-5 years (parameter agegr = c(0, 1, 3, 5, 12, 24, 36, 48, 60)/12). These are aggregated to a cumulative hazards over the age group 0-4 years and converted to probabilities to estimate $_{5}q_0$.

Add a binary indicator whether a death occurred and a date of death variable, placed 0.5 months in the month the death occurred.

```
br <- lapply(br, function(x){x$death <- x$b5 == "no"; x})
br <- lapply(br, function(x){x$dod <- x$b3 + x$b7 + 0.5; x})
```

Calculate $_{5}q_0$ for period 0-4, 5-9, and 10-14 years preceding the survey.

```
u5mr <- lapply(br, calc_nqx, by=~SurveyId+CountryName+SurveyYear, strata=NULL)
u5mr <- do.call(rbind, u5mr)
```

Merge DHS StatCompiler indicators

Identify the indicator IDs associated with TFR, ASFR 15-19, $_{35}q_{15}$, and $_{5}q_0$.

```
indic <- dhs_indicators()

indic[grepl("TFR 15-49", ShortName), .(IndicatorId, ShortName, Label)]
```

```
##      IndicatorId ShortName                   Label
## 1: FE_FRTR_W_TFR TFR 15-49 Total fertility rate 15-49
```

```
indic[grepl("ASFR 15-19", ShortName), .(IndicatorId, ShortName, Label)]
```

```
##      IndicatorId  ShortName
## 1: FE_FRTR_W_A15 ASFR 15-19
## 2: FE_FRTT_W_A15 ASFR 15-19
##                                                    Label
## 1:                   Age specific fertility rate: 15-19
## 2: Age specific fertility rate: 15-19 (five year periods)
```

```
indic[grepl("Probability of dying", ShortName), .(IndicatorId, Definition)]
```

```
##      IndicatorId
## 1: MM_AMPB_W_AMP
## 2: MM_AMPB_M_AMP
##                                                    Definition
```

```
## 1: Probability of dying between exact age 15 and 50 (35q15) for women
## 2:   Probability of dying between exact age 15 and 50 (35q15) for men
```

```r
indic[grepl("Under-five mortality", ShortName), .(IndicatorId, Label)]
```

```
##        IndicatorId                    Label
## 1: CM_ECMR_C_U5M Under-five mortality rate
```

Query estimates from DHS API and merge with calculated estimates.

```r
tfr_dhs <- dhs_data(indicatorIds = "FE_FRTR_W_TFR",
                    surveyId = tfr$SurveyId)
tfr <- merge(tfr, tfr_dhs[ , .(SurveyId, Value)])

asfr15to19_dhs <- dhs_data(indicatorIds = "FE_FRTR_W_A15",
                           surveyId = asfr15to19$SurveyId)
asfr15to19 <- merge(asfr15to19, asfr15to19_dhs[ , .(SurveyId, Value)])

q3515_dhs <- dhs_data(indicatorIds = c("MM_AMPB_W_AMP", "MM_AMPB_M_AMP"),
                      surveyId = q3515$SurveyId)
q3515_dhs$mm1 <- c(MM_AMPB_M_AMP = "male", MM_AMPB_W_AMP = "female")[q3515_dhs$IndicatorId]
q3515 <- merge(q3515, q3515_dhs[ , .(SurveyId, mm1, Value)])

u5mr_dhs <- dhs_data(indicatorIds = "CM_ECMT_C_U5M",
                     surveyYearStart = 2005,
                     breakdown = "all")[SurveyId %in% u5mr$SurveyId]
u5mr_dhs$tips <- u5mr_dhs$CharacteristicLabel
u5mr <- merge(u5mr, u5mr_dhs[, .(SurveyId, tips, Value)])
```

View estimates

Table 1: TFR

SurveyId	CountryName	SurveyYear	tips	tfr	se_tfr	Value
AO2015DHS	Angola	2015	0-2	6.2	0.139	6.2
BF2010DHS	Burkina Faso	2010	0-2	6.0	0.099	6.0
BJ2006DHS	Benin	2006	0-2	5.7	0.072	5.7
BJ2012DHS	Benin	2012	0-2	4.9	0.066	4.9
BU2010DHS	Burundi	2010	0-2	6.4	0.098	6.4
BU2016DHS	Burundi	2016	0-2	5.5	0.076	5.5

Table 2: ASFR 15-19

SurveyId	CountryName	SurveyYear	agegr	tips	asfr	se_asfr	Value
AO2015DHS	Angola	2015	15-19	0-2	0.163	0.0073	163
BF2010DHS	Burkina Faso	2010	15-19	0-2	0.130	0.0049	130
BJ2006DHS	Benin	2006	15-19	0-2	0.112	0.0045	112
BJ2012DHS	Benin	2012	15-19	0-2	0.094	0.0038	94
BU2010DHS	Burundi	2010	15-19	0-2	0.065	0.0040	65
BU2016DHS	Burundi	2016	15-19	0-2	0.058	0.0030	58

Table 3: 35q15

SurveyId	mm1	CountryName	SurveyYear	tips	est	se	ci_l	ci_u	Value
AO2015DHS	female	Angola	2015	0-6	0.110	0.010	0.090	0.130	110
AO2015DHS	male	Angola	2015	0-6	0.182	0.013	0.157	0.207	182
BF2010DHS	female	Burkina Faso	2010	0-6	0.146	0.008	0.131	0.161	146
BF2010DHS	male	Burkina Faso	2010	0-6	0.145	0.007	0.130	0.158	145
BJ2006DHS	female	Benin	2006	0-6	0.127	0.006	0.114	0.139	127
BJ2006DHS	male	Benin	2006	0-6	0.161	0.008	0.146	0.177	162

Table 4: 5q0

	SurveyId	tips	CountryName	SurveyYear	est	se	ci_l	ci_u	Value
1	AO2015DHS	0-4	Angola	2015	0.066	0.004	0.058	0.073	68
3	AO2015DHS	5-9	Angola	2015	0.093	0.005	0.082	0.103	95
2	AO2015DHS	10-14	Angola	2015	0.146	0.008	0.130	0.163	145
4	BF2010DHS	0-4	Burkina Faso	2010	0.121	0.004	0.113	0.129	129
6	BF2010DHS	5-9	Burkina Faso	2010	0.169	0.005	0.160	0.179	168
5	BF2010DHS	10-14	Burkina Faso	2010	0.180	0.006	0.169	0.191	177

Check that TFR, ASFR, and $_{35}q_{15}$ estimates exactly match

Estimates for fertility rates and adult mortality rates should exactly match those produced as standard DHS indicators.

```
## TFR matches exactly
with(tfr, table(round(tfr, 1) == Value))

##
## TRUE
##   66
```

```
## ASFR 15-19 matches exactly
with(asfr15to19, table(round(1000*asfr) == Value))

##
## TRUE
##   66
```

```
## 35q15 matches exactly for >80%
with(q3515, table(round(1000*est) == Value))

##
## FALSE   TRUE
##    23     87
```

```
with(q3515, table(round(1000*est) - Value))

##
## -8 -3 -1  0  1  5
##  1  1 16 87  4  1
```

```
subset(q3515, abs((round(1000*est) - Value)) > 1)

##     SurveyId   mm1 CountryName SurveyYear tips      est       se
```

```
## 57 MZ2011DHS female  Mozambique       2011  0-6 0.1914953 0.01021433
## 58 MZ2011DHS   male  Mozambique       2011  0-6 0.2383925 0.01186020
## 97 UG2011DHS female      Uganda       2011  0-6 0.2062956 0.01222652
##          ci_l      ci_u Value
## 57 0.1712256 0.2112692   199
## 58 0.2147886 0.2612869   241
## 97 0.1819667 0.2299010   201
```

Compare $_5q_0$ estimates

```r
u5mr$tips <- factor(u5mr$tips, c("0-4", "5-9", "10-14"))
ggplot(u5mr, aes(1000*est, Value, color=tips)) +
  geom_abline(slope=1, color="grey") +
  geom_point() +
  coord_fixed() +
  xlab("demogsurv::calc_nqx()") +
  ylab("DHS StatCompiler") +
  ggtitle("5q0 comparison")
```