Exclusive Breastfeeding in Ethiopia: Unified Reproducible Analysis

2024-25 Ethiopia Demographic and Health Survey

Dr. Aynalem Adugna

July 18, 2026

Contents

1 How to use this report

This document unifies the thirty instructional code modules into one executable R Markdown workflow. Edit the `kr_file` parameter in the YAML header so it points to the Ethiopia DHS 2024-25 Kid Recode (`.dta`) file, then select **Knit > Knit to PDF** in RStudio. The report installs missing R packages, reconstructs and validates the exclusive breastfeeding indicator, performs the survey-weighted analyses, prints all major tables, creates all analytical figures, generates the epidemiologic DAG and programmatic conceptual framework, and exports reusable CSV and PNG files.

The DHS data file is not distributed with this report. Access must remain governed by the DHS Program data-use agreement.

2 Module 1: Package management and reproducibility

```r
required_packages <- c(
  "haven", "dplyr", "tidyr", "purrr", "stringr", "forcats",
  "survey", "broom", "emmeans", "ggplot2", "knitr", "scales"
)

missing_packages <- required_packages[
  !required_packages %in% rownames(installed.packages())
]

if (length(missing_packages) > 0L) {
  install.packages(missing_packages, dependencies = TRUE)
}

invisible(lapply(required_packages, library, character.only = TRUE))

set.seed(2026)

output_dir <- normalizePath(
  params$output_dir,
  winslash = "/",
  mustWork = FALSE
)

dir.create(output_dir, recursive = TRUE, showWarnings = FALSE)
dir.create(file.path(output_dir, "tables"), recursive = TRUE, showWarnings = FALSE)
dir.create(file.path(output_dir, "figures"), recursive = TRUE, showWarnings = FALSE)

session_start <- Sys.time()
```

3 Module 2: Read the DHS Kid Recode file

```r
kr_file <- params$kr_file

if (!file.exists(kr_file)) {
  stop(
    "The KR data file was not found. Edit params$kr_file in the YAML header.\nCurrent path: ",
    kr_file
  )
}

kr2024 <- haven::read_dta(kr_file)

file_summary <- data.frame(
  Item = c("Data file", "Records", "Variables"),
  Value = c(kr_file, format(nrow(kr2024), big.mark = ","), ncol(kr2024))
)

knitr::kable(file_summary, caption = "Input-data summary", booktabs = TRUE)
```

Table 1: Input-data summary

Item	Value
Data file	C:/Users/aynal/OneDrive/Documents/ETH-DHS-2024-25/2024DATA/ETKR8ADT/ETKR8AFL.dta
Records	13,565
Variables	1301

4 Module 3: Confirm required variables

```
essential_variables <- c(
  "caseid", "v005", "v021", "v022", "v024", "v025", "v106", "v190",
  "v012", "v201", "b5", "b9", "b19", "bidx", "b4", "m4", "v409"
)

missing_essential <- setdiff(essential_variables, names(kr2024))

if (length(missing_essential) > 0L) {
  stop(
    "Required variables are missing from the KR file: ",
    paste(missing_essential, collapse = ", ")
  )
}

variable_audit <- data.frame(
  Variable = essential_variables,
  Present = essential_variables %in% names(kr2024)
)

knitr::kable(variable_audit, caption = "Required-variable audit", booktabs = TRUE)
```

Table 2: Required-variable audit

Variable	Present
caseid	TRUE
v005	TRUE
v021	TRUE
v022	TRUE
v024	TRUE
v025	TRUE
v106	TRUE
v190	TRUE
v012	TRUE
v201	TRUE
b5	TRUE
b9	TRUE
b19	TRUE
bidx	TRUE
b4	TRUE
m4	TRUE
v409	TRUE

5 Module 4: Identify DHS-8 liquid and food variables

```r
v413_variables <- grep("^v413", names(kr2024), value = TRUE)
v414_variables <- grep("^v414", names(kr2024), value = TRUE)

feeding_variable_inventory <- data.frame(
  Series = c(rep("v413 liquid series", length(v413_variables)),
             rep("v414 food series", length(v414_variables))),
  Variable = c(v413_variables, v414_variables)
)

knitr::kable(
  feeding_variable_inventory,
  caption = "DHS-8 liquid and food variables detected in the KR file",
  booktabs = TRUE,
  longtable = TRUE
)
```

Table 3: DHS-8 liquid and food variables detected in the KR file

Series	Variable
v413 liquid series	v413
v413 liquid series	v413s
v413 liquid series	v413a
v413 liquid series	v413as
v413 liquid series	v413b
v413 liquid series	v413bs
v413 liquid series	v413c
v413 liquid series	v413d
v413 liquid series	v413e
v413 liquid series	v413f
v413 liquid series	v413g
v413 liquid series	v413h
v413 liquid series	v413i
v414 food series	v414a
v414 food series	v414b
v414 food series	v414c
v414 food series	v414d
v414 food series	v414e
v414 food series	v414f
v414 food series	v414g
v414 food series	v414h
v414 food series	v414i
v414 food series	v414j
v414 food series	v414k
v414 food series	v414l
v414 food series	v414m
v414 food series	v414n
v414 food series	v414o
v414 food series	v414p
v414 food series	v414q
v414 food series	v414r
v414 food series	v414s
v414 food series	v414t
v414 food series	v414u

Series	Variable
v414 food series	v414v
v414 food series	v414w
v414 food series	v414wa
v414 food series	v414wb
v414 food series	v414wc
v414 food series	v414wd
v414 food series	v414we

6 Module 5: Define helper functions

```
consumed_1_to_7 <- function(x) {
  as.integer(!is.na(x) & x >= 1 & x <= 7)
}

any_consumed <- function(data, variables) {
  variables <- intersect(variables, names(data))
  if (length(variables) == 0L) return(rep(0L, nrow(data)))
  consumption_matrix <- vapply(
    data[variables], consumed_1_to_7, integer(nrow(data))
  )
  if (is.null(dim(consumption_matrix))) {
    consumption_matrix <- matrix(consumption_matrix, ncol = 1L)
  }
  as.integer(rowSums(consumption_matrix, na.rm = TRUE) > 0)
}

format_p <- function(p) {
  ifelse(is.na(p), NA_character_, ifelse(p < 0.001, "<0.001", sprintf("%.3f", p)))
}

report_table <- function(x, caption, digits = 2) {
  numeric_columns <- vapply(x, is.numeric, logical(1))
  x[numeric_columns] <- lapply(x[numeric_columns], round, digits = digits)
  knitr::kable(x, caption = caption, booktabs = TRUE, longtable = TRUE, row.names = FALSE)
}
```

7 Module 6: Define and document the analytical sample

```
sample_flow <- data.frame(
  Step = c(
    "All KR records",
    "Living children",
    "Living infants aged 0-5 completed months",
    "Most recent birth",
    "Infant resides with interviewed mother"
  ),
  Unweighted_N = c(
    nrow(kr2024),
    sum(kr2024$b5 == 1, na.rm = TRUE),
```

```
    sum(kr2024$b5 == 1 & kr2024$b19 >= 0 & kr2024$b19 < 6, na.rm = TRUE),
    sum(kr2024$b5 == 1 & kr2024$b19 >= 0 & kr2024$b19 < 6 & kr2024$bidx == 1, na.rm = TRUE),
    sum(kr2024$b5 == 1 & kr2024$b19 >= 0 & kr2024$b19 < 6 & kr2024$bidx == 1 & kr2024$b9 == 0, na.rm = TRUE
  )
)

report_table(sample_flow, "Analytical sample flow")
```

Table 4: Analytical sample flow

Step	Unweighted_N
All KR records	13565
Living children	12921
Living infants aged 0-5 completed months	1364
Most recent birth	1342
Infant resides with interviewed mother	1325

```
write.csv(sample_flow, file.path(output_dir, "tables", "Table_Sample_Flow.csv"), row.names = FALSE)
```

```
ebf_data <- kr2024 %>%
  filter(
    b5 == 1,
    b19 >= 0,
    b19 < 6,
    bidx == 1,
    b9 == 0
  ) %>%
  mutate(weight = v005 / 1000000)

age_sample_distribution <- ebf_data %>%
  count(b19, name = "Unweighted_N") %>%
  rename(Infant_Age_Completed_Months = b19)

report_table(age_sample_distribution, "Eligible sample by completed month of infant age")
```

Table 5: Eligible sample by completed month of infant age

Infant_Age_Completed_Months	Unweighted_N
0	207
1	233
2	237
3	213
4	224
5	211

8 Module 7: Reconstruct current breastfeeding

```
ebf_data <- ebf_data %>%
  mutate(currently_breastfeeding = as.integer(m4 == 95))

current_bf_check <- ebf_data %>%
  count(currently_breastfeeding, name = "Unweighted_N") %>%
```

```r
  mutate(Status = recode(as.character(currently_breastfeeding),
                         `0` = "Not currently breastfeeding",
                         `1` = "Currently breastfeeding")) %>%
  select(Status, Unweighted_N)

report_table(current_bf_check, "Current breastfeeding reconstructed from m4 = 95")
```

Table 6: Current breastfeeding reconstructed from m4 = 95

Status	Unweighted_N
Not currently breastfeeding	119
Currently breastfeeding	1206

9 Module 8: Reconstruct water, liquids, milk, and solid-food indicators

```r
base_liquid_variables <- c("v409a", "v410", "v410a", "v412c")
all_liquid_variables <- unique(c(base_liquid_variables, v413_variables))
milk_variables <- c("v411", "v411a")
solid_variables <- unique(c(v414_variables, "v412a", "v412b", "m39a"))

ebf_data <- ebf_data %>%
  mutate(
    water = consumed_1_to_7(v409),
    liquids = any_consumed(cur_data(), all_liquid_variables),
    milk = any_consumed(cur_data(), milk_variables),
    solids = any_consumed(cur_data(), solid_variables)
  )

feeding_component_check <- bind_rows(
  ebf_data %>% count(water) %>% mutate(Component = "Plain water", Response = water),
  ebf_data %>% count(liquids) %>% mutate(Component = "Other nonmilk liquids", Response = liquids),
  ebf_data %>% count(milk) %>% mutate(Component = "Other milk or formula", Response = milk),
  ebf_data %>% count(solids) %>% mutate(Component = "Solid, semisolid, or soft food", Response = solids)
) %>%
  transmute(Component, Response = ifelse(Response == 1, "Yes", "No"), Unweighted_N = n)

report_table(feeding_component_check, "Feeding-component diagnostics")
```

Table 7: Feeding-component diagnostics

Component	Response	Unweighted_N
Plain water	No	973
Plain water	Yes	352
Other nonmilk liquids	No	1230
Other nonmilk liquids	Yes	95
Other milk or formula	No	1138
Other milk or formula	Yes	187
Solid, semisolid, or soft food	No	1143
Solid, semisolid, or soft food	Yes	182

10 Module 9: Construct the DHS breastfeeding-status hierarchy and EBF outcome

```
ebf_data <- ebf_data %>%
  mutate(
    breastfeeding_status = 1L,
    breastfeeding_status = ifelse(water == 1, 2L, breastfeeding_status),
    breastfeeding_status = ifelse(liquids == 1, 3L, breastfeeding_status),
    breastfeeding_status = ifelse(milk == 1, 4L, breastfeeding_status),
    breastfeeding_status = ifelse(solids == 1, 5L, breastfeeding_status),
    breastfeeding_status = ifelse(currently_breastfeeding == 0, 0L, breastfeeding_status),
    ebf = as.integer(breastfeeding_status == 1L)
  )

status_labels <- c(
  `0` = "Not breastfeeding",
  `1` = "Exclusively breastfeeding",
  `2` = "Breastfeeding plus plain water",
  `3` = "Breastfeeding plus nonmilk liquids",
  `4` = "Breastfeeding plus other milk/formula",
  `5` = "Breastfeeding plus complementary foods"
)

feeding_status_table <- ebf_data %>%
  count(breastfeeding_status, name = "Unweighted_N") %>%
  mutate(Status = unname(status_labels[as.character(breastfeeding_status)])) %>%
  select(Status, Unweighted_N)

report_table(feeding_status_table, "DHS breastfeeding-status hierarchy")
```

Table 8: DHS breastfeeding-status hierarchy

Status	Unweighted_N
Not breastfeeding	119
Exclusively breastfeeding	733
Breastfeeding plus plain water	161
Breastfeeding plus nonmilk liquids	30
Breastfeeding plus other milk/formula	125
Breastfeeding plus complementary foods	157

11 Module 10: Validate the reconstructed EBF indicator

```
unweighted_ebf <- ebf_data %>%
  summarise(
    Unweighted_N = n(),
    EBF_N = sum(ebf == 1, na.rm = TRUE),
    Unweighted_EBF_Percent = 100 * mean(ebf, na.rm = TRUE)
  )

report_table(unweighted_ebf, "Unweighted EBF diagnostic")
```

Table 9: Unweighted EBF diagnostic

Unweighted_N	EBF_N	Unweighted_EBF_Percent
1325	733	55.32

12 Module 11: Create the final analysis variables

```
analysis_data <- ebf_data %>%
  mutate(
    ebf_factor = factor(ebf, levels = c(0, 1),
                        labels = c("Not exclusively breastfed", "Exclusively breastfed")),
    infant_age_month = factor(b19, levels = 0:5, labels = paste0(0:5, " months")),
    infant_sex = factor(b4, levels = c(1, 2), labels = c("Male", "Female")),
    residence = factor(v025, levels = c(2, 1), labels = c("Rural", "Urban")),
    maternal_education = factor(v106, levels = c(0, 1, 2, 3),
                                labels = c("No education", "Primary", "Secondary", "Higher")),
    wealth_quintile = factor(v190, levels = 1:5,
                             labels = c("Poorest", "Poorer", "Middle", "Richer", "Richest")),
    wealth_score = ifelse(v190 %in% 1:5, as.numeric(v190), NA_real_),
    maternal_age_group = cut(v012, breaks = c(14, 24, 34, 49),
                             labels = c("15-24", "25-34", "35-49"), right = TRUE),
    maternal_age_years = ifelse(v012 >= 15 & v012 <= 49, as.numeric(v012), NA_real_),
    parity_group = case_when(
      v201 == 1 ~ "1",
      v201 %in% 2:3 ~ "2-3",
      v201 %in% 4:5 ~ "4-5",
      v201 >= 6 & v201 < 90 ~ "6+",
      TRUE ~ NA_character_
    ),
    parity_group = factor(parity_group, levels = c("1", "2-3", "4-5", "6+")),
    region = case_when(
      as.numeric(v024) == 1  ~ "Tigray",
      as.numeric(v024) == 2  ~ "Afar",
      as.numeric(v024) == 3  ~ "Amhara",
      as.numeric(v024) == 4  ~ "Oromia",
      as.numeric(v024) == 5  ~ "Somali",
      as.numeric(v024) == 6  ~ "Benishangul-Gumuz",
      as.numeric(v024) == 7  ~ "Central Ethiopia",
      as.numeric(v024) == 8  ~ "Sidama",
      as.numeric(v024) == 9  ~ "Southwest Ethiopia",
      as.numeric(v024) == 10 ~ "South Ethiopia",
      as.numeric(v024) == 12 ~ "Gambella",
      as.numeric(v024) == 13 ~ "Harari",
      as.numeric(v024) == 14 ~ "Addis Ababa",
      as.numeric(v024) == 15 ~ "Dire Dawa",
      TRUE ~ NA_character_
    ),
    region = factor(region, levels = c(
      "Oromia", "Tigray", "Afar", "Amhara", "Somali", "Benishangul-Gumuz",
      "Central Ethiopia", "Sidama", "Southwest Ethiopia", "South Ethiopia",
      "Gambella", "Harari", "Addis Ababa", "Dire Dawa"
    )),
    harmonized_region = case_when(
```

```r
      as.numeric(v024) == 1  ~ "Tigray",
      as.numeric(v024) == 2  ~ "Afar",
      as.numeric(v024) == 3  ~ "Amhara",
      as.numeric(v024) == 4  ~ "Oromia",
      as.numeric(v024) == 5  ~ "Somali",
      as.numeric(v024) == 6  ~ "Benishangul-Gumuz",
      as.numeric(v024) %in% c(7, 8, 9, 10) ~ "Former SNNPR",
      as.numeric(v024) == 12 ~ "Gambella",
      as.numeric(v024) == 13 ~ "Harari",
      as.numeric(v024) == 14 ~ "Addis Ababa",
      as.numeric(v024) == 15 ~ "Dire Dawa",
      TRUE ~ NA_character_
    ),
    harmonized_region = factor(harmonized_region, levels = c(
      "Oromia", "Tigray", "Afar", "Amhara", "Somali", "Benishangul-Gumuz",
      "Former SNNPR", "Gambella", "Harari", "Addis Ababa", "Dire Dawa"
    ))
  )
```

13 Module 12: Audit constructed variables

```r
variables_to_check <- c(
  "ebf_factor", "infant_age_month", "infant_sex", "residence",
  "maternal_education", "wealth_quintile", "maternal_age_group",
  "parity_group", "region", "harmonized_region"
)

constructed_variable_audit <- map_dfr(variables_to_check, function(variable_name) {
  analysis_data %>%
    count(.data[[variable_name]], name = "Unweighted_N", .drop = FALSE) %>%
    transmute(
      Variable = variable_name,
      Category = as.character(.data[[variable_name]]),
      Unweighted_N
    )
})

report_table(constructed_variable_audit, "Constructed-variable audit")
```

Table 10: Constructed-variable audit

Variable	Category	Unweighted_N
ebf_factor	Not exclusively breastfed	592
ebf_factor	Exclusively breastfed	733
infant_age_month	0 months	207
infant_age_month	1 months	233
infant_age_month	2 months	237
infant_age_month	3 months	213
infant_age_month	4 months	224
infant_age_month	5 months	211
infant_sex	Male	659
infant_sex	Female	666
residence	Rural	933

Variable	Category	Unweighted_N
residence	Urban	392
maternal_education	No education	496
maternal_education	Primary	496
maternal_education	Secondary	201
maternal_education	Higher	132
wealth_quintile	Poorest	467
wealth_quintile	Poorer	228
wealth_quintile	Middle	196
wealth_quintile	Richer	177
wealth_quintile	Richest	257
maternal_age_group	15-24	410
maternal_age_group	25-34	672
maternal_age_group	35-49	243
parity_group	1	281
parity_group	2-3	516
parity_group	4-5	253
parity_group	6+	275
region	Oromia	131
region	Tigray	123
region	Afar	115
region	Amhara	106
region	Somali	123
region	Benishangul-Gumuz	82
region	Central Ethiopia	86
region	Sidama	76
region	Southwest Ethiopia	76
region	South Ethiopia	147
region	Gambella	69
region	Harari	72
region	Addis Ababa	48
region	Dire Dawa	71
harmonized_region	Oromia	131
harmonized_region	Tigray	123
harmonized_region	Afar	115
harmonized_region	Amhara	106
harmonized_region	Somali	123
harmonized_region	Benishangul-Gumuz	82
harmonized_region	Former SNNPR	385
harmonized_region	Gambella	69
harmonized_region	Harari	72
harmonized_region	Addis Ababa	48
harmonized_region	Dire Dawa	71

```r
write.csv(constructed_variable_audit,
          file.path(output_dir, "tables", "Table_Constructed_Variable_Audit.csv"),
          row.names = FALSE)
```

14 Module 13: DAG-guided covariate specification

The epidemiologic model treats infant age and region as primary explanatory dimensions and retains infant sex, residence, maternal education, household wealth, maternal age, and parity as prespecified covariates. Selection is based on causal and substantive reasoning rather than automated significance screening.

15 Module 14: Build the complex survey design

```
design <- svydesign(
  ids = ~v021,
  strata = ~v022,
  weights = ~weight,
  data = analysis_data,
  nest = TRUE
)

survey_design_summary <- data.frame(
  Measure = c("Eligible infants", "Sampled PSUs", "Sampling strata", "Sum of normalized weights"),
  Value = c(
    nrow(analysis_data),
    dplyr::n_distinct(analysis_data$v021),
    dplyr::n_distinct(analysis_data$v022),
    round(sum(analysis_data$weight, na.rm = TRUE), 2)
  )
)

report_table(survey_design_summary, "Complex survey-design summary")
```

Table 11: Complex survey-design summary

Measure	Value
Eligible infants	1325.00
Sampled PSUs	619.00
Sampling strata	27.00
Sum of normalized weights	1318.45

16 Module 15: Define the reusable weighted-prevalence function

```
weighted_ebf_table <- function(design_object, data_object, grouping_variable, grouping_label) {
  grouping_formula <- as.formula(paste0("~", grouping_variable))

  estimate <- svyby(
    formula = ~ebf,
    by = grouping_formula,
    design = design_object,
    FUN = svymean,
    na.rm = TRUE,
    vartype = c("se", "ci"),
    level = 0.95,
    keep.names = FALSE,
    drop.empty.groups = FALSE
  ) %>% as.data.frame()

  names(estimate)[1] <- grouping_variable

  unweighted <- data_object %>%
    filter(!is.na(.data[[grouping_variable]]), !is.na(ebf)) %>%
    count(.data[[grouping_variable]], name = "Unweighted_N", .drop = FALSE)
```

```
  names(unweighted)[1] <- grouping_variable

  estimate %>%
    left_join(unweighted, by = grouping_variable) %>%
    transmute(
      Characteristic = grouping_label,
      Category = as.character(.data[[grouping_variable]]),
      Unweighted_N,
      EBF_Percent = 100 * ebf,
      Standard_Error_Percentage_Points = 100 * se,
      Lower_95_CI = 100 * ci_l,
      Upper_95_CI = 100 * ci_u
    )
}
```

17 Module 16: Estimate and validate national EBF prevalence

```
national_object <- svymean(~ebf, design, na.rm = TRUE)
national_ci <- confint(national_object, level = 0.95)

national_results <- data.frame(
  Characteristic = "National",
  Category = "Ethiopia",
  Unweighted_N = nrow(analysis_data),
  EBF_Percent = 100 * coef(national_object)[1],
  Standard_Error_Percentage_Points = 100 * SE(national_object)[1],
  Lower_95_CI = 100 * national_ci[1, 1],
  Upper_95_CI = 100 * national_ci[1, 2]
)

published_comparison <- national_results %>%
  transmute(
    Published_Percent = params$published_ebf_percent,
    Reconstructed_Percent = EBF_Percent,
    Absolute_Difference_Percentage_Points = abs(EBF_Percent - params$published_ebf_percent),
    Tolerance_Percentage_Points = params$validation_tolerance_points,
    Validation = ifelse(
      Absolute_Difference_Percentage_Points <= params$validation_tolerance_points,
      "Pass: close agreement",
      "Review required"
    )
  )

report_table(national_results, "National survey-weighted exclusive breastfeeding prevalence")
```

Table 12: National survey-weighted exclusive breastfeeding prevalence

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
National	Ethiopia	1325	58.07	2.51	53.15	63

```
report_table(published_comparison, "Validation against the published EDHS estimate")
```

Table 13: Validation against the published EDHS estimate

Published_Percent	Reconstructed_Percent	Absolute_Difference_Percentage_Points	Tolerance_Percentage_Points	Validation
57.3	58.07	0.77	1	Pass: close agreement

18 Module 17: Estimate the age-specific EBF pattern

```
age_results <- weighted_ebf_table(
  design, analysis_data, "infant_age_month", "Infant age"
) %>%
  mutate(Infant_Age_Months = as.integer(stringr::str_extract(Category, "^[0-5]")))

report_table(age_results, "Exclusive breastfeeding by completed month of infant age")
```

Table 14: Exclusive breastfeeding by completed month of infant age

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI	Infant_Age_Months
Infant age	0 months	207	85.22	4.75	75.90	94.53	0
Infant age	1 months	233	69.46	5.19	59.28	79.64	1
Infant age	2 months	237	63.40	5.58	52.46	74.34	2
Infant age	3 months	213	71.26	4.95	61.56	80.97	3
Infant age	4 months	224	42.08	5.21	31.88	52.29	4
Infant age	5 months	211	26.19	5.84	14.74	37.64	5

19 Module 18: Estimate EBF by infant sex

```
sex_results <- weighted_ebf_table(
  design, analysis_data, "infant_sex", "Infant sex"
)
report_table(sex_results, "Exclusive breastfeeding by infant sex")
```

Table 15: Exclusive breastfeeding by infant sex

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Infant sex	Male	659	60.15	3.35	53.58	66.71
Infant sex	Female	666	56.09	3.36	49.50	62.69

20 Module 19: Estimate EBF by residence

```r
residence_results <- weighted_ebf_table(
  design, analysis_data, "residence", "Residence"
)
report_table(residence_results, "Exclusive breastfeeding by urban-rural residence")
```

Table 16: Exclusive breastfeeding by urban-rural residence

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Residence	Rural	933	59.79	3.23	53.45	66.12
Residence	Urban	392	54.43	3.75	47.07	61.78

21 Module 20: Estimate EBF by maternal education

```r
education_results <- weighted_ebf_table(
  design, analysis_data, "maternal_education", "Maternal education"
)
report_table(education_results, "Exclusive breastfeeding by maternal education")
```

Table 17: Exclusive breastfeeding by maternal education

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Maternal education	No education	496	55.76	3.91	48.09	63.44
Maternal education	Primary	496	58.99	3.49	52.15	65.83
Maternal education	Secondary	201	58.51	6.03	46.70	70.33
Maternal education	Higher	132	61.42	7.11	47.48	75.36

22 Module 21: Estimate EBF by household wealth

```r
wealth_results <- weighted_ebf_table(
  design, analysis_data, "wealth_quintile", "Household wealth"
)
report_table(wealth_results, "Exclusive breastfeeding by household wealth quintile")
```

Table 18: Exclusive breastfeeding by household wealth quintile

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Household wealth	Poorest	467	55.28	4.01	47.42	63.14
Household wealth	Poorer	228	54.72	5.03	44.86	64.59
Household wealth	Middle	196	57.16	5.63	46.13	68.19
Household wealth	Richer	177	64.96	6.14	52.93	77.00

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Household wealth	Richest	257	57.46	4.79	48.08	66.84

23 Module 22: Estimate EBF by maternal age

```
maternal_age_results <- weighted_ebf_table(
  design, analysis_data, "maternal_age_group", "Maternal age"
)
report_table(maternal_age_results, "Exclusive breastfeeding by maternal age")
```

Table 19: Exclusive breastfeeding by maternal age

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Maternal age	15-24	410	55.97	4.20	47.75	64.20
Maternal age	25-34	672	57.63	3.48	50.82	64.44
Maternal age	35-49	243	61.95	5.10	51.95	71.95

24 Module 23: Estimate EBF by parity

```
parity_results <- weighted_ebf_table(
  design, analysis_data, "parity_group", "Parity"
)
report_table(parity_results, "Exclusive breastfeeding by parity")
```

Table 20: Exclusive breastfeeding by parity

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Parity	1	281	65.33	4.55	56.42	74.24
Parity	2-3	516	53.80	3.64	46.66	60.94
Parity	4-5	253	62.46	4.83	52.99	71.94
Parity	6+	275	55.65	5.10	45.66	65.64

25 Module 24: Estimate current and harmonized regional prevalence

```
region_results <- weighted_ebf_table(
  design, analysis_data, "region", "Current region"
)

harmonized_region_results <- weighted_ebf_table(
  design, analysis_data, "harmonized_region", "Harmonized region"
)

report_table(region_results, "Exclusive breastfeeding by current region")
```

Table 21: Exclusive breastfeeding by current region

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Current region	Oromia	131	48.73	5.06	38.80	58.65
Current region	Tigray	123	77.94	3.28	71.51	84.37
Current region	Afar	115	41.09	5.71	29.90	52.28
Current region	Amhara	106	72.26	5.22	62.04	82.49
Current region	Somali	123	28.88	5.07	18.95	38.80
Current region	Benishangul-Gumuz	82	70.72	5.33	60.27	81.16
Current region	Central Ethiopia	86	58.09	5.94	46.46	69.73
Current region	Sidama	76	59.58	5.60	48.61	70.55
Current region	Southwest Ethiopia	76	65.54	6.03	53.73	77.35
Current region	South Ethiopia	147	58.38	5.06	48.47	68.29
Current region	Gambella	69	42.14	7.12	28.19	56.10
Current region	Harari	72	51.62	6.73	38.43	64.81
Current region	Addis Ababa	48	47.88	7.73	32.74	63.03
Current region	Dire Dawa	71	49.95	7.62	35.01	64.89

```
report_table(harmonized_region_results, "Exclusive breastfeeding by harmonized 2016-compatible region")
```

Table 22: Exclusive breastfeeding by harmonized 2016-compatible region

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Harmonized region	Oromia	131	48.73	5.06	38.80	58.65
Harmonized region	Tigray	123	77.94	3.28	71.51	84.37
Harmonized region	Afar	115	41.09	5.71	29.90	52.28
Harmonized region	Amhara	106	72.26	5.22	62.04	82.49
Harmonized region	Somali	123	28.88	5.07	18.95	38.80
Harmonized region	Benishangul-Gumuz	82	70.72	5.33	60.27	81.16
Harmonized region	Former SNNPR	385	59.37	3.12	53.25	65.49
Harmonized region	Gambella	69	42.14	7.12	28.19	56.10
Harmonized region	Harari	72	51.62	6.73	38.43	64.81

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Harmonized region	Addis Ababa	48	47.88	7.73	32.74	63.03
Harmonized region	Dire Dawa	71	49.95	7.62	35.01	64.89

26 Module 25: Combine all descriptive results

```r
all_prevalence_results <- bind_rows(
  national_results,
  age_results %>% select(-Infant_Age_Months),
  sex_results,
  residence_results,
  education_results,
  wealth_results,
  maternal_age_results,
  parity_results,
  region_results
) %>%
  mutate(
    Estimate_95_CI = sprintf("%.1f%% (%.1f-%.1f)", EBF_Percent, Lower_95_CI, Upper_95_CI)
  )

report_table(all_prevalence_results, "Combined survey-weighted descriptive results")
```

Table 23: Combined survey-weighted descriptive results

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI	Estimate_95_CI
National	Ethiopia	1325	58.07	2.51	53.15	63.00	58.1% (53.1-63.0)
Infant age	0 months	207	85.22	4.75	75.90	94.53	85.2% (75.9-94.5)
Infant age	1 months	233	69.46	5.19	59.28	79.64	69.5% (59.3-79.6)
Infant age	2 months	237	63.40	5.58	52.46	74.34	63.4% (52.5-74.3)
Infant age	3 months	213	71.26	4.95	61.56	80.97	71.3% (61.6-81.0)
Infant age	4 months	224	42.08	5.21	31.88	52.29	42.1% (31.9-52.3)
Infant age	5 months	211	26.19	5.84	14.74	37.64	26.2% (14.7-37.6)
Infant sex	Male	659	60.15	3.35	53.58	66.71	60.1% (53.6-66.7)
Infant sex	Female	666	56.09	3.36	49.50	62.69	56.1% (49.5-62.7)
Residence	Rural	933	59.79	3.23	53.45	66.12	59.8% (53.4-66.1)
Residence	Urban	392	54.43	3.75	47.07	61.78	54.4% (47.1-61.8)
Maternal education	No education	496	55.76	3.91	48.09	63.44	55.8% (48.1-63.4)

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage	Lower_95_CI	Upper_95_CI	Estimate_95_CI
Maternal education	Primary	496	58.99	3.49	52.15	65.83	59.0% (52.2-65.8)
Maternal education	Secondary	201	58.51	6.03	46.70	70.33	58.5% (46.7-70.3)
Maternal education	Higher	132	61.42	7.11	47.48	75.36	61.4% (47.5-75.4)
Household wealth	Poorest	467	55.28	4.01	47.42	63.14	55.3% (47.4-63.1)
Household wealth	Poorer	228	54.72	5.03	44.86	64.59	54.7% (44.9-64.6)
Household wealth	Middle	196	57.16	5.63	46.13	68.19	57.2% (46.1-68.2)
Household wealth	Richer	177	64.96	6.14	52.93	77.00	65.0% (52.9-77.0)
Household wealth	Richest	257	57.46	4.79	48.08	66.84	57.5% (48.1-66.8)
Maternal age	15-24	410	55.97	4.20	47.75	64.20	56.0% (47.7-64.2)
Maternal age	25-34	672	57.63	3.48	50.82	64.44	57.6% (50.8-64.4)
Maternal age	35-49	243	61.95	5.10	51.95	71.95	62.0% (52.0-72.0)
Parity	1	281	65.33	4.55	56.42	74.24	65.3% (56.4-74.2)
Parity	2-3	516	53.80	3.64	46.66	60.94	53.8% (46.7-60.9)
Parity	4-5	253	62.46	4.83	52.99	71.94	62.5% (53.0-71.9)
Parity	6+	275	55.65	5.10	45.66	65.64	55.6% (45.7-65.6)
Current region	Oromia	131	48.73	5.06	38.80	58.65	48.7% (38.8-58.6)
Current region	Tigray	123	77.94	3.28	71.51	84.37	77.9% (71.5-84.4)
Current region	Afar	115	41.09	5.71	29.90	52.28	41.1% (29.9-52.3)
Current region	Amhara	106	72.26	5.22	62.04	82.49	72.3% (62.0-82.5)
Current region	Somali	123	28.88	5.07	18.95	38.80	28.9% (18.9-38.8)
Current region	Benishangul-Gumuz	82	70.72	5.33	60.27	81.16	70.7% (60.3-81.2)
Current region	Central Ethiopia	86	58.09	5.94	46.46	69.73	58.1% (46.5-69.7)
Current region	Sidama	76	59.58	5.60	48.61	70.55	59.6% (48.6-70.6)
Current region	Southwest Ethiopia	76	65.54	6.03	53.73	77.35	65.5% (53.7-77.4)
Current region	South Ethiopia	147	58.38	5.06	48.47	68.29	58.4% (48.5-68.3)
Current region	Gambella	69	42.14	7.12	28.19	56.10	42.1% (28.2-56.1)
Current region	Harari	72	51.62	6.73	38.43	64.81	51.6% (38.4-64.8)

Characteristic	Category	Unweighted_N	EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI	Estimate_95_CI
Current region	Addis Ababa	48	47.88	7.73	32.74	63.03	47.9% (32.7-63.0)
Current region	Dire Dawa	71	49.95	7.62	35.01	64.89	49.9% (35.0-64.9)

```r
write.csv(all_prevalence_results,
          file.path(output_dir, "tables", "Table_All_Weighted_Descriptive_Results.csv"),
          row.names = FALSE)
```

27 Module 26: Design-adjusted bivariate tests

```r
test_variables <- c(
  infant_age_month = "Infant age",
  infant_sex = "Infant sex",
  residence = "Residence",
  maternal_education = "Maternal education",
  wealth_quintile = "Household wealth",
  maternal_age_group = "Maternal age",
  parity_group = "Parity",
  region = "Region"
)

rao_scott_results <- imap_dfr(test_variables, function(label, variable) {
  test <- svychisq(
    as.formula(paste0("~ebf_factor + ", variable)),
    design = design,
    statistic = "F"
  )
  data.frame(
    Predictor = label,
    Design_Adjusted_F = unname(test$statistic),
    Numerator_df = unname(test$parameter[1]),
    Denominator_df = unname(test$parameter[2]),
    P_Value = unname(test$p.value),
    P_Value_Formatted = format_p(unname(test$p.value))
  )
})

report_table(rao_scott_results, "Rao-Scott design-adjusted tests of association")
```

Table 24: Rao-Scott design-adjusted tests of association

Predictor	Design_Adjusted_F	Numerator_df	Denominator_df	P_Value	P_Value_Formatted
Infant age	13.73	4.53	2683.60	0.00	<0.001
Infant sex	0.82	1.00	592.00	0.37	0.365
Residence	1.17	1.00	592.00	0.28	0.279
Maternal education	0.23	2.97	1756.38	0.88	0.877
Household wealth	0.66	3.58	2118.16	0.60	0.601
Maternal age	0.43	1.96	1158.41	0.64	0.644
Parity	1.74	2.90	1718.61	0.16	0.159
Region	6.72	4.76	2818.22	0.00	<0.001

```
write.csv(rao_scott_results,
          file.path(output_dir, "tables", "Table_Rao_Scott_Tests.csv"),
          row.names = FALSE)
```

28 Module 27: Survey-weighted logistic regression

```
extract_or_table <- function(fitted_model, model_name) {
  broom::tidy(fitted_model, conf.int = TRUE, exponentiate = TRUE) %>%
    filter(term != "(Intercept)") %>%
    transmute(
      Model = model_name,
      Term = term,
      Odds_Ratio = estimate,
      Lower_95_CI = conf.low,
      Upper_95_CI = conf.high,
      P_Value = p.value,
      P_Value_Formatted = format_p(p.value)
    )
}

unadjusted_predictors <- c(
  "infant_age_month", "infant_sex", "residence", "maternal_education",
  "wealth_quintile", "maternal_age_group", "parity_group", "region"
)

unadjusted_models <- setNames(
  lapply(unadjusted_predictors, function(variable) {
    svyglm(as.formula(paste0("ebf ~ ", variable)),
           design = design, family = quasibinomial())
  }),
  unadjusted_predictors
)

unadjusted_or_table <- imap_dfr(
  unadjusted_models,
  ~extract_or_table(.x, paste("Unadjusted:", .y))
)

unadjusted_global_tests <- imap_dfr(unadjusted_models, function(model, variable) {
  test <- regTermTest(model, as.formula(paste0("~", variable)))
  data.frame(
    Predictor = variable,
    Wald_F = unname(test$Ftest),
    Numerator_df = unname(test$df),
    Denominator_df = unname(test$ddf),
    P_Value = unname(test$p),
    P_Value_Formatted = format_p(unname(test$p))
  )
})

adjusted_model <- svyglm(
  ebf ~ infant_age_month + infant_sex + residence + maternal_education +
    wealth_quintile + maternal_age_group + parity_group + region,
  design = design,
```

```r
  family = quasibinomial()
)

adjusted_or_table <- extract_or_table(adjusted_model, "Fully adjusted model")

adjusted_terms <- unadjusted_predictors
adjusted_global_tests <- map_dfr(adjusted_terms, function(variable) {
  test <- regTermTest(adjusted_model, as.formula(paste0("~", variable)))
  data.frame(
    Predictor = variable,
    Wald_F = unname(test$Ftest),
    Numerator_df = unname(test$df),
    Denominator_df = unname(test$ddf),
    P_Value = unname(test$p),
    P_Value_Formatted = format_p(unname(test$p))
  )
})
```

```r
report_table(unadjusted_or_table, "Unadjusted survey-weighted logistic-regression estimates")
```

Table 25: Unadjusted survey-weighted logistic-regression estimates

Model	Term	Odds_Ratio	Lower_95_CI	Upper_95_CI	P_Value	P_Value_Formatted
Unadjusted: infant_age_month	infant_age_month1 months	0.39	0.16	0.94	0.04	0.037
Unadjusted: infant_age_month	infant_age_month2 months	0.30	0.12	0.74	0.01	0.009
Unadjusted: infant_age_month	infant_age_month3 months	0.43	0.22	0.86	0.02	0.016
Unadjusted: infant_age_month	infant_age_month4 months	0.13	0.05	0.30	0.00	<0.001
Unadjusted: infant_age_month	infant_age_month5 months	0.06	0.02	0.16	0.00	<0.001
Unadjusted: infant_sex	infant_sexFemale	0.85	0.59	1.21	0.37	0.365
Unadjusted: residence	residenceUrban	0.80	0.54	1.20	0.28	0.280
Unadjusted: maternal_education	maternal_educationPrimary	1.14	0.77	1.69	0.51	0.511
Unadjusted: maternal_education	maternal_educationSecondary	1.12	0.64	1.97	0.70	0.696
Unadjusted: maternal_education	maternal_educationHigher	1.26	0.65	2.46	0.49	0.493
Unadjusted: wealth_quintile	wealth_quintilePoorer	0.98	0.62	1.54	0.92	0.922
Unadjusted: wealth_quintile	wealth_quintileMiddle	1.08	0.63	1.83	0.78	0.778
Unadjusted: wealth_quintile	wealth_quintileRicher	1.50	0.80	2.82	0.21	0.207
Unadjusted: wealth_quintile	wealth_quintileRichest	1.09	0.66	1.80	0.73	0.728
Unadjusted: maternal_age_group	maternal_age_group25-34	1.07	0.70	1.63	0.75	0.753
Unadjusted: maternal_age_group	maternal_age_group35-49	1.28	0.77	2.12	0.33	0.335
Unadjusted: parity_group	parity_group2-3	0.62	0.39	0.98	0.04	0.041

Model	Term	Odds_Ratio	Lower_95_CI	Upper_95_CI	P_Value	P_Value_Formatted
Unadjusted: parity_group	parity_group4-5	0.88	0.50	1.56	0.67	0.669
Unadjusted: parity_group	parity_group6+	0.67	0.39	1.15	0.14	0.142
Unadjusted: region	regionTigray	3.72	2.15	6.42	0.00	<0.001
Unadjusted: region	regionAfar	0.73	0.40	1.35	0.32	0.320
Unadjusted: region	regionAmhara	2.74	1.43	5.24	0.00	0.002
Unadjusted: region	regionSomali	0.43	0.23	0.80	0.01	0.008
Unadjusted: region	regionBenishangul-Gumuz	2.54	1.34	4.84	0.00	0.005
Unadjusted: region	regionCentral Ethiopia	1.46	0.78	2.72	0.23	0.234
Unadjusted: region	regionSidama	1.55	0.85	2.84	0.16	0.155
Unadjusted: region	regionSouthwest Ethiopia	2.00	1.04	3.87	0.04	0.039
Unadjusted: region	regionSouth Ethiopia	1.48	0.83	2.61	0.18	0.181
Unadjusted: region	regionGambella	0.77	0.38	1.54	0.45	0.455
Unadjusted: region	regionHarari	1.12	0.58	2.18	0.73	0.731
Unadjusted: region	regionAddis Ababa	0.97	0.47	2.00	0.93	0.928
Unadjusted: region	regionDire Dawa	1.05	0.51	2.16	0.89	0.894

```
report_table(unadjusted_global_tests, "Global Wald tests for unadjusted models")
```

Table 26: Global Wald tests for unadjusted models

Predictor	Wald_F	Numerator_df	Denominator_df	P_Value	P_Value_Formatted
infant_age_month	10.46	5	587	0.00	<0.001
infant_sex	0.82	1	591	0.37	0.365
residence	1.17	1	591	0.28	0.280
maternal_education	0.22	3	589	0.88	0.879
wealth_quintile	0.48	4	588	0.75	0.753
maternal_age_group	0.47	2	590	0.63	0.625
parity_group	2.05	3	589	0.11	0.106
region	6.23	13	579	0.00	<0.001

```
report_table(adjusted_or_table, "Adjusted survey-weighted logistic-regression estimates")
```

Table 27: Adjusted survey-weighted logistic-regression estimates

Model	Term	Odds_Ratio	Lower_95_CI	Upper_95_CI	P_Value	P_Value_Formatted
Fully adjusted model	infant_age_month1 months	0.36	0.15	0.90	0.03	0.028
Fully adjusted model	infant_age_month2 months	0.27	0.11	0.67	0.01	0.005
Fully adjusted model	infant_age_month3 months	0.40	0.20	0.82	0.01	0.012
Fully adjusted model	infant_age_month4 months	0.09	0.04	0.20	0.00	<0.001
Fully adjusted model	infant_age_month5 months	0.05	0.02	0.12	0.00	<0.001
Fully adjusted model	infant_sexFemale	0.96	0.63	1.47	0.86	0.857
Fully adjusted model	residenceUrban	0.54	0.23	1.30	0.17	0.171

Model	Term	Odds_Ratio	Lower_95_CI	Upper_95_CI	P_Value	P_Value_Formatted
Fully adjusted model	maternal_educationPrimary	1.19	0.73	1.95	0.49	0.490
Fully adjusted model	maternal_educationSecondary	0.83	0.42	1.63	0.59	0.589
Fully adjusted model	maternal_educationHigher	0.79	0.35	1.78	0.57	0.574
Fully adjusted model	wealth_quintilePoorer	1.04	0.57	1.87	0.90	0.903
Fully adjusted model	wealth_quintileMiddle	1.28	0.63	2.60	0.50	0.498
Fully adjusted model	wealth_quintileRicher	1.93	0.87	4.29	0.10	0.104
Fully adjusted model	wealth_quintileRichest	2.55	0.76	8.55	0.13	0.129
Fully adjusted model	maternal_age_group25-34	1.24	0.70	2.21	0.46	0.457
Fully adjusted model	maternal_age_group35-49	1.31	0.56	3.05	0.53	0.532
Fully adjusted model	parity_group2-3	0.56	0.30	1.05	0.07	0.073
Fully adjusted model	parity_group4-5	0.78	0.32	1.92	0.59	0.592
Fully adjusted model	parity_group6+	0.56	0.23	1.34	0.19	0.192
Fully adjusted model	regionTigray	4.79	2.42	9.50	0.00	<0.001
Fully adjusted model	regionAfar	0.67	0.30	1.53	0.34	0.342
Fully adjusted model	regionAmhara	2.70	1.29	5.65	0.01	0.008
Fully adjusted model	regionSomali	0.33	0.15	0.74	0.01	0.007
Fully adjusted model	regionBenishangul-Gumuz	2.36	1.06	5.22	0.03	0.035
Fully adjusted model	regionCentral Ethiopia	1.69	0.76	3.74	0.20	0.198
Fully adjusted model	regionSidama	1.40	0.68	2.90	0.36	0.364
Fully adjusted model	regionSouthwest Ethiopia	2.33	1.10	4.91	0.03	0.027
Fully adjusted model	regionSouth Ethiopia	1.59	0.87	2.91	0.13	0.133
Fully adjusted model	regionGambella	0.72	0.30	1.71	0.45	0.450
Fully adjusted model	regionHarari	0.85	0.43	1.66	0.63	0.628
Fully adjusted model	regionAddis Ababa	0.86	0.27	2.70	0.79	0.790
Fully adjusted model	regionDire Dawa	0.74	0.30	1.83	0.51	0.513

```r
report_table(adjusted_global_tests, "Global Wald tests for the adjusted model")
```

Table 28: Global Wald tests for the adjusted model

Predictor	Wald_F	Numerator_df	Denominator_df	P_Value	P_Value_Formatted
infant_age_month	13.24	5	560	0.00	<0.001
infant_sex	0.03	1	560	0.86	0.857
residence	1.88	1	560	0.17	0.171
maternal_education	0.79	3	560	0.50	0.498
wealth_quintile	1.08	4	560	0.36	0.364
maternal_age_group	0.29	2	560	0.75	0.750
parity_group	1.63	3	560	0.18	0.181
region	5.17	13	560	0.00	<0.001

```
write.csv(unadjusted_or_table, file.path(output_dir, "tables", "Table_Unadjusted_Odds_Ratios.csv"), row.nam
write.csv(adjusted_or_table, file.path(output_dir, "tables", "Table_Adjusted_Odds_Ratios.csv"), row.names =
write.csv(adjusted_global_tests, file.path(output_dir, "tables", "Table_Adjusted_Global_Wald_Tests.csv"), r
```

29 Module 28: Adjusted odds ratios and trend models

```
age_trend_model <- svyglm(
  ebf ~ b19 + infant_sex + residence + maternal_education + wealth_quintile +
    maternal_age_group + parity_group + region,
  design = design,
  family = quasibinomial()
)

age_trend_result <- broom::tidy(age_trend_model, conf.int = TRUE, exponentiate = TRUE) %>%
  filter(term == "b19") %>%
  transmute(
    Predictor = "Each additional completed month of infant age",
    Adjusted_Odds_Ratio = estimate,
    Lower_95_CI = conf.low,
    Upper_95_CI = conf.high,
    P_Value = p.value,
    P_Value_Formatted = format_p(p.value)
  )

wealth_trend_model <- svyglm(
  ebf ~ infant_age_month + infant_sex + residence + maternal_education + wealth_score +
    maternal_age_group + parity_group + region,
  design = design,
  family = quasibinomial()
)

wealth_trend_result <- broom::tidy(wealth_trend_model, conf.int = TRUE, exponentiate = TRUE) %>%
  filter(term == "wealth_score") %>%
  transmute(
    Predictor = "One-quintile increase in household wealth",
    Adjusted_Odds_Ratio = estimate,
    Lower_95_CI = conf.low,
    Upper_95_CI = conf.high,
    P_Value = p.value,
    P_Value_Formatted = format_p(p.value)
  )
```

```r
report_table(age_trend_result, "Adjusted linear trend in EBF by infant age")
```

Table 29: Adjusted linear trend in EBF by infant age

Predictor	Adjusted_Odds_Ratio	Lower_95_CI	Upper_95_CI	P_Value	P_Value_Formatted
Each additional completed month of infant age	0.59	0.51	0.68	0	<0.001

```r
report_table(wealth_trend_result, "Adjusted linear trend in EBF by household wealth")
```

Table 30: Adjusted linear trend in EBF by household wealth

Predictor	Adjusted_Odds_Ratio	Lower_95_CI	Upper_95_CI	P_Value	P_Value_Formatted
One-quintile increase in household wealth	1.25	0.99	1.58	0.07	0.066

```r
harmonized_model <- svyglm(
  ebf ~ infant_age_month + infant_sex + residence + maternal_education +
    wealth_quintile + maternal_age_group + parity_group + harmonized_region,
  design = design,
  family = quasibinomial()
)

harmonized_model_results <- extract_or_table(
  harmonized_model,
  "Adjusted model using harmonized regions"
)

model_variables <- c(
  "ebf", "infant_age_month", "infant_sex", "residence", "maternal_education",
  "wealth_quintile", "maternal_age_group", "parity_group", "region"
)

model_sample_flow <- data.frame(
  Stage = c("Eligible analytical sample", "Complete cases for primary adjusted model", "Excluded for missin
  Unweighted_N = c(
    nrow(analysis_data),
    sum(complete.cases(analysis_data[model_variables])),
    nrow(analysis_data) - sum(complete.cases(analysis_data[model_variables]))
  )
)

report_table(model_sample_flow, "Model sample flow")
```

Table 31: Model sample flow

Stage	Unweighted_N
Eligible analytical sample	1325
Complete cases for primary adjusted model	1325
Excluded for missing model covariates	0

```r
report_table(harmonized_model_results, "Sensitivity model using harmonized regions")
```

Table 32: Sensitivity model using harmonized regions

Model	Term	Odds_Ratio	Lower_95_CI	Upper_95_CI	P_Value	P_Value_Formatted
Adjusted model using harmonized regions	infant_age_month1 months	0.36	0.15	0.89	0.03	0.028
Adjusted model using harmonized regions	infant_age_month2 months	0.27	0.11	0.67	0.00	0.005
Adjusted model using harmonized regions	infant_age_month3 months	0.40	0.20	0.82	0.01	0.012
Adjusted model using harmonized regions	infant_age_month4 months	0.09	0.04	0.20	0.00	<0.001
Adjusted model using harmonized regions	infant_age_month5 months	0.05	0.02	0.12	0.00	<0.001
Adjusted model using harmonized regions	infant_sexFemale	0.96	0.63	1.46	0.84	0.841
Adjusted model using harmonized regions	residenceUrban	0.55	0.23	1.31	0.18	0.177
Adjusted model using harmonized regions	maternal_educationPrimary	1.18	0.72	1.92	0.52	0.518
Adjusted model using harmonized regions	maternal_educationSecondary	0.82	0.42	1.61	0.56	0.565
Adjusted model using harmonized regions	maternal_educationHigher	0.79	0.35	1.76	0.56	0.559
Adjusted model using harmonized regions	wealth_quintilePoorer	1.03	0.57	1.86	0.92	0.916
Adjusted model using harmonized regions	wealth_quintileMiddle	1.27	0.63	2.58	0.50	0.500
Adjusted model using harmonized regions	wealth_quintileRicher	1.92	0.87	4.23	0.11	0.105
Adjusted model using harmonized regions	wealth_quintileRichest	2.49	0.75	8.28	0.14	0.136
Adjusted model using harmonized regions	maternal_age_group25-34	1.25	0.71	2.21	0.44	0.438
Adjusted model using harmonized regions	maternal_age_group35-49	1.33	0.57	3.09	0.50	0.504
Adjusted model using harmonized regions	parity_group2-3	0.56	0.30	1.05	0.07	0.072
Adjusted model using harmonized regions	parity_group4-5	0.78	0.32	1.89	0.58	0.579
Adjusted model using harmonized regions	parity_group6+	0.55	0.23	1.30	0.17	0.172
Adjusted model using harmonized regions	harmonized_regionTigray	4.79	2.42	9.49	0.00	<0.001
Adjusted model using harmonized regions	harmonized_regionAfar	0.67	0.30	1.52	0.34	0.336
Adjusted model using harmonized regions	harmonized_regionAmhara	2.69	1.29	5.63	0.01	0.009
Adjusted model using harmonized regions	harmonized_regionSomali	0.33	0.15	0.74	0.01	0.007
Adjusted model using harmonized regions	harmonized_regionBenishangul-Gumuz	2.35	1.06	5.20	0.04	0.036
Adjusted model using harmonized regions	harmonized_regionFormer SNNPR	1.64	0.95	2.84	0.07	0.075
Adjusted model using harmonized regions	harmonized_regionGambella	0.72	0.30	1.71	0.45	0.451
Adjusted model using harmonized regions	harmonized_regionHarari	0.85	0.43	1.67	0.64	0.637

Model	Term	Odds_Ratio	Lower_95_CI	Upper_95_CI	P_Value	P_Value_Formatted
Adjusted model using harmonized regions	harmonized_regionAddis Ababa	0.86	0.27	2.71	0.80	0.797
Adjusted model using harmonized regions	harmonized_regionDire Dawa	0.74	0.30	1.84	0.52	0.518

```
write.csv(model_sample_flow, file.path(output_dir, "tables", "Table_Model_Sample_Flow.csv"), row.names = FA
write.csv(harmonized_model_results, file.path(output_dir, "tables", "Table_Harmonized_Region_Model.csv"), r
```

30 Module 29: Adjusted predicted probabilities

```
extract_emmeans_probability <- function(emm_object, grouping_name, output_name) {
  emm_df <- as.data.frame(emm_object)
  probability_name <- intersect(c("prob", "response"), names(emm_df))[1]
  lower_name <- intersect(c("asymp.LCL", "lower.CL"), names(emm_df))[1]
  upper_name <- intersect(c("asymp.UCL", "upper.CL"), names(emm_df))[1]

  if (any(is.na(c(probability_name, lower_name, upper_name)))) {
    stop("Expected probability or confidence-limit columns were not found in emmeans output.")
  }

  emm_df %>%
    transmute(
      Category = as.character(.data[[grouping_name]]),
      Adjusted_EBF_Percent = 100 * .data[[probability_name]],
      Standard_Error_Percentage_Points = 100 * SE,
      Lower_95_CI = 100 * .data[[lower_name]],
      Upper_95_CI = 100 * .data[[upper_name]]
    ) %>%
    rename(!!output_name := Category)
}
```

```
age_nuisance_factors <- c(
  "infant_sex", "residence", "maternal_education", "wealth_quintile",
  "maternal_age_group", "parity_group", "region"
)

age_emmeans <- emmeans(
  adjusted_model,
  specs = ~infant_age_month,
  nuisance = age_nuisance_factors,
  wt.nuis = "proportional",
  weights = "proportional",
  type = "response"
)

adjusted_age_probabilities <- extract_emmeans_probability(
  age_emmeans, "infant_age_month", "Infant_Age"
) %>%
  mutate(Infant_Age_Months = as.integer(str_extract(Infant_Age, "^[0-5]")))

region_nuisance_factors <- c(
  "infant_age_month", "infant_sex", "residence", "maternal_education",
```

```r
  "wealth_quintile", "maternal_age_group", "parity_group"
)

region_emmeans <- emmeans(
  adjusted_model,
  specs = ~region,
  nuisance = region_nuisance_factors,
  wt.nuis = "proportional",
  weights = "proportional",
  type = "response"
)

adjusted_region_probabilities <- extract_emmeans_probability(
  region_emmeans, "region", "Region"
)

report_table(adjusted_age_probabilities, "Adjusted EBF probabilities by infant age")
```

Table 33: Adjusted EBF probabilities by infant age

Infant_Age	Adjusted_EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI	Infant_Age_Months
0 months	84.77	4.42	74.00	91.58	0
1 months	66.92	5.74	54.90	77.07	1
2 months	59.91	5.78	48.25	70.55	2
3 months	69.18	4.79	59.10	77.71	3
4 months	32.29	4.68	23.86	42.05	4
5 months	21.14	4.93	13.05	32.38	5

```r
report_table(adjusted_region_probabilities, "Adjusted EBF probabilities by region")
```

Table 34: Adjusted EBF probabilities by region

Region	Adjusted_EBF_Percent	Standard_Error_Percentage_Points	Lower_95_CI	Upper_95_CI
Oromia	50.31	6.32	38.15	62.43
Tigray	82.92	3.42	75.16	88.62
Afar	40.51	7.66	26.75	55.95
Amhara	73.21	5.88	60.29	83.10
Somali	25.29	5.86	15.56	38.35
Benishangul-Gumuz	70.46	6.81	55.66	81.92
Central Ethiopia	63.07	7.76	47.05	76.65
Sidama	58.65	6.70	45.22	70.90
Southwest Ethiopia	70.21	5.96	57.42	80.46
South Ethiopia	61.69	4.57	52.43	70.18
Gambella	42.01	8.60	26.62	59.13
Harari	46.15	6.61	33.73	59.08
Addis Ababa	46.41	13.42	23.12	71.38
Dire Dawa	42.80	9.79	25.48	62.10

```r
write.csv(adjusted_age_probabilities, file.path(output_dir, "tables", "Table_Adjusted_Age_Probabilities.csv
write.csv(adjusted_region_probabilities, file.path(output_dir, "tables", "Table_Adjusted_Region_Probabiliti
```

31 Module 30: Publication figures, DAG, conceptual framework, and export

31.1 Figure 1. Epidemiologic DAG

```r
dag_boxes <- data.frame(
  label = c("Regional context", "Residence", "Maternal education", "Household wealth",
            "Maternal age", "Parity", "Infant sex", "Infant age", "Exclusive breastfeeding"),
  x = c(5, 2, 5, 8, 1.5, 3.8, 6.2, 8.5, 5),
  y = c(9, 7, 7, 7, 4.7, 4.7, 4.7, 4.7, 1.7),
  width = c(4.2, 2.3, 2.5, 2.3, 1.8, 1.8, 1.8, 1.8, 3.4),
  height = c(1.0, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 1.1),
  group = c("context", rep("socioeconomic", 3), rep("proximal", 4), "outcome")
)

dag_edges <- data.frame(
  from = c(1,1,1,2,3,2,2,3,3,4,4,4,5,6,7,8),
  to   = c(2,3,4,3,4,5,6,5,6,6,7,8,9,9,9,9)
)

get_anchor <- function(index) dag_boxes[index, ]

dag_plot <- ggplot() +
  geom_segment(data = dag_edges, aes(
    x = dag_boxes$x[from], y = dag_boxes$y[from] - dag_boxes$height[from]/2,
    xend = dag_boxes$x[to], yend = dag_boxes$y[to] + dag_boxes$height[to]/2
  ), arrow = arrow(length = unit(0.16, "inches")), linewidth = 0.45) +
  geom_rect(data = dag_boxes, aes(
    xmin = x - width/2, xmax = x + width/2,
    ymin = y - height/2, ymax = y + height/2,
    fill = group
  ), color = "black", linewidth = 0.45) +
  geom_text(data = dag_boxes, aes(x = x, y = y, label = label),
            fontface = "bold", size = 3.4, lineheight = 0.95) +
  scale_fill_manual(values = c(
    context = "#D9EAF7", socioeconomic = "#DDEFD9",
    proximal = "#F7E7C6", outcome = "#F5D5DD"
  )) +
  coord_cartesian(xlim = c(0, 10), ylim = c(0.5, 10), clip = "off") +
  labs(title = "Directed Acyclic Graph of Hypothesized Determinants of EBF",
       subtitle = "Infants aged 0-5 completed months, Ethiopia DHS 2024-25") +
  theme_void(base_size = 11) +
  theme(legend.position = "none", plot.title = element_text(hjust = 0.5, face = "bold"),
        plot.subtitle = element_text(hjust = 0.5))

dag_plot
```

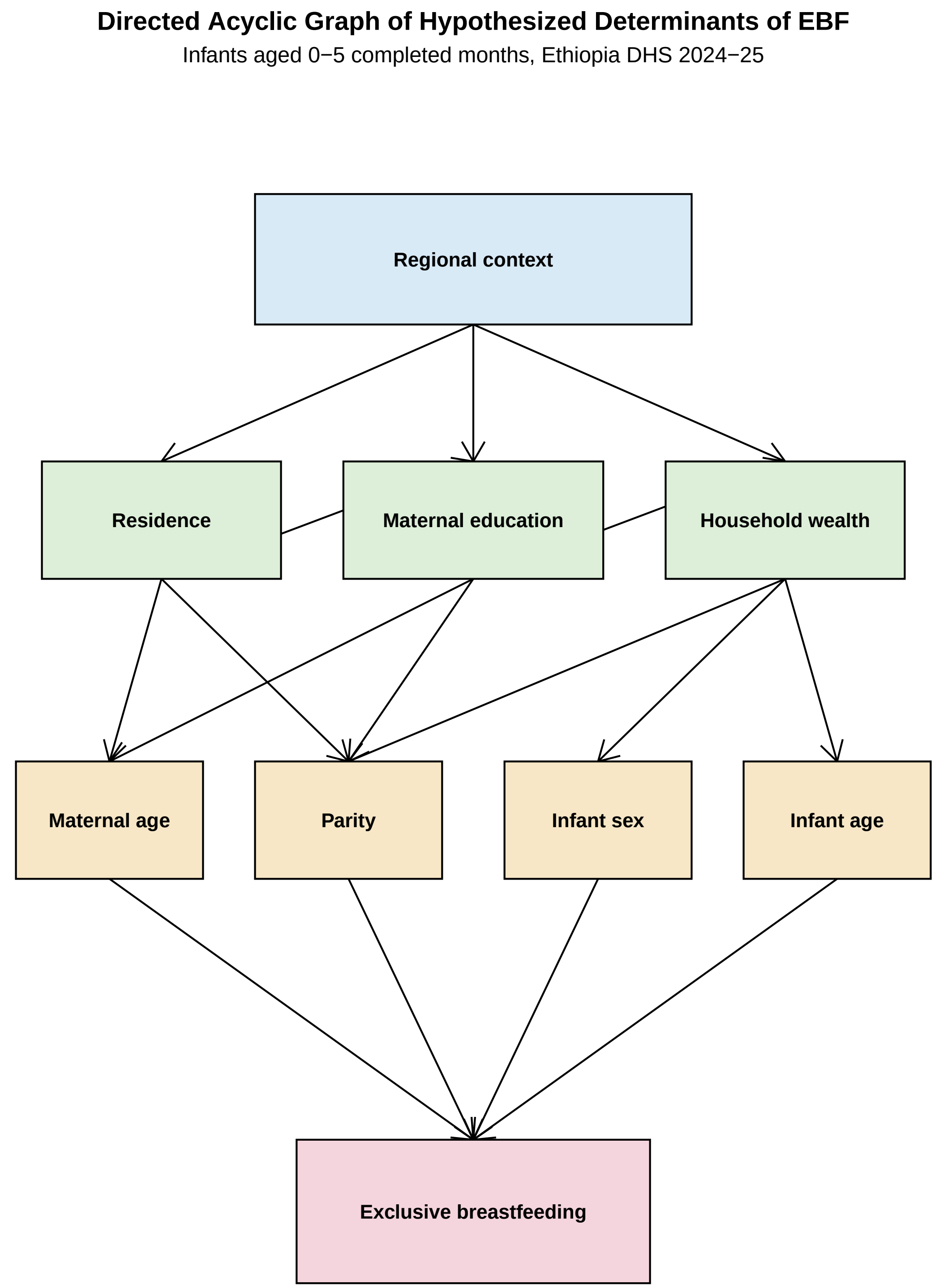

Figure 1: Directed acyclic graph guiding covariate selection for the exclusive breastfeeding analysis.

```
ggsave(file.path(output_dir, "figures", "Figure_1_EBF_DAG.png"), dag_plot,
       width = 8, height = 10, dpi = 300)
```

31.2 Figure 2. Programmatic conceptual framework

```
cf_boxes <- data.frame(
  label = c(
    "POLICIES & HEALTH SYSTEM\nIYCF guidance, maternity protection,\nBaby-Friendly services, HEWs",
    "HIGH INITIATION\nBreastfeeding begins at or soon after birth",
    "CONTINUATION CHALLENGES\nPerceived insufficient milk; family influence;\nworkload; early liquids/foods
    "ENABLING CONDITIONS\nMaternal confidence; family support;\ncommunity norms; repeated skilled counselin
    "SUSTAINED EBF TO 6 MONTHS\nImproved infant, maternal, and societal outcomes",
    "STRATEGIC PRIORITY\nIntensify support during months 2-5\nand tailor implementation to regional context
  ),
  x = c(1.8, 5, 8.2, 5, 8.2, 5),
  y = c(7.2, 7.2, 7.2, 4.6, 4.6, 1.8),
  width = c(3.0, 2.7, 3.0, 4.6, 3.0, 5.4),
  height = c(1.5, 1.5, 1.8, 1.6, 1.5, 1.5),
  type = c("input", "continuum", "barrier", "enabler", "outcome", "priority")
)

cf_edges <- data.frame(from = c(1,2,3,4,4,5), to = c(2,3,4,5,6,6))

cf_plot <- ggplot() +
  geom_segment(data = cf_edges, aes(
    x = cf_boxes$x[from], y = cf_boxes$y[from],
    xend = cf_boxes$x[to], yend = cf_boxes$y[to]
  ), arrow = arrow(length = unit(0.16, "inches")), linewidth = 0.55) +
  geom_rect(data = cf_boxes, aes(
    xmin = x - width/2, xmax = x + width/2,
    ymin = y - height/2, ymax = y + height/2,
    fill = type
  ), color = "black", linewidth = 0.5) +
  geom_text(data = cf_boxes, aes(x = x, y = y, label = label),
            size = 3.2, lineheight = 0.95) +
  scale_fill_manual(values = c(
    input = "#D9EAF7", continuum = "#DDEFD9", barrier = "#F7E7C6",
    enabler = "#E8DDF2", outcome = "#D8EEDB", priority = "#DCE7F7"
  )) +
  coord_cartesian(xlim = c(0, 10), ylim = c(0.5, 8.5), clip = "off") +
  labs(title = "From Initiation to Continuation",
       subtitle = "A programmatic conceptual framework for exclusive breastfeeding in Ethiopia") +
  theme_void(base_size = 11) +
  theme(legend.position = "none", plot.title = element_text(hjust = 0.5, face = "bold"),
        plot.subtitle = element_text(hjust = 0.5))

cf_plot
```

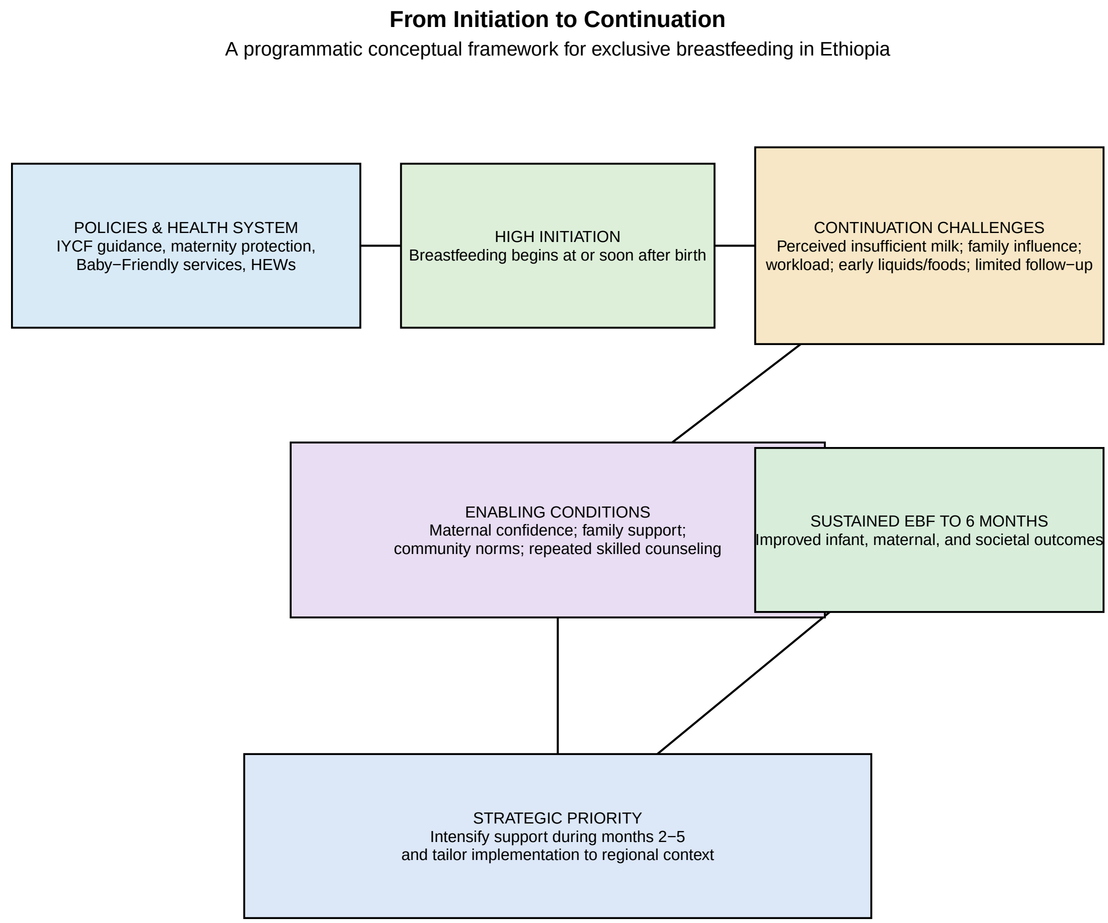

Figure 2: Programmatic conceptual framework for sustaining exclusive breastfeeding through six months.

```
ggsave(file.path(output_dir, "figures", "Figure_2_EBF_Conceptual_Framework.png"), cf_plot,
       width = 10, height = 8, dpi = 300)
```

31.3 Figure 3. Observed and adjusted EBF by infant age

```
age_plot_data <- age_results %>%
  select(Infant_Age_Months, Observed = EBF_Percent, Observed_Lower = Lower_95_CI, Observed_Upper = Upper_95
  left_join(
    adjusted_age_probabilities %>%
      select(Infant_Age_Months, Adjusted = Adjusted_EBF_Percent,
             Adjusted_Lower = Lower_95_CI, Adjusted_Upper = Upper_95_CI),
    by = "Infant_Age_Months"
  )

age_plot_long <- bind_rows(
  age_plot_data %>% transmute(Infant_Age_Months, Series = "Observed prevalence", Estimate = Observed,
                              Lower = Observed_Lower, Upper = Observed_Upper),
  age_plot_data %>% transmute(Infant_Age_Months, Series = "Adjusted probability", Estimate = Adjusted,
                              Lower = Adjusted_Lower, Upper = Adjusted_Upper)
```

```
)

age_figure <- ggplot(age_plot_long, aes(x = Infant_Age_Months, y = Estimate,
                                        group = Series, linetype = Series, shape = Series)) +
  geom_line(linewidth = 0.8) +
  geom_point(size = 2.4) +
  geom_errorbar(aes(ymin = Lower, ymax = Upper), width = 0.08, linewidth = 0.45) +
  scale_x_continuous(breaks = 0:5) +
  scale_y_continuous(limits = c(0, 100), breaks = seq(0, 100, 20), labels = scales::label_percent(scale = 1
  labs(x = "Infant age in completed months", y = "Exclusive breastfeeding",
       title = "Age-Specific Exclusive Breastfeeding in Ethiopia",
       linetype = NULL, shape = NULL) +
  theme_minimal(base_size = 11) +
  theme(legend.position = "bottom")

age_figure
```

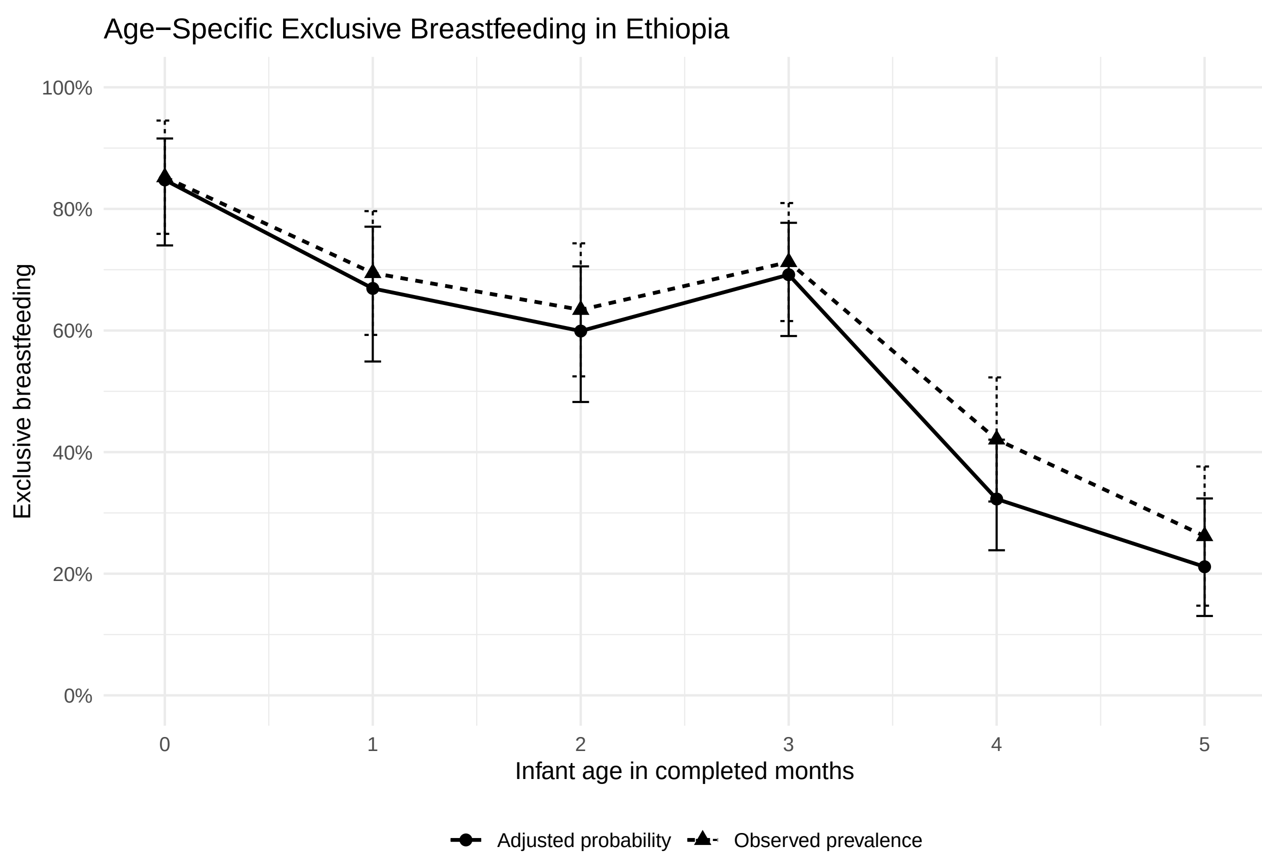

Figure 3: Observed survey-weighted prevalence and adjusted predicted probability of exclusive breastfeeding by infant age.

```
ggsave(file.path(output_dir, "figures", "Figure_3_EBF_By_Infant_Age.png"), age_figure,
       width = 8, height = 5.5, dpi = 300)
```

31.4 Figure 4. Adjusted EBF probability by region

```r
regional_figure <- adjusted_region_probabilities %>%
  mutate(Region = forcats::fct_reorder(Region, Adjusted_EBF_Percent)) %>%
  ggplot(aes(x = Adjusted_EBF_Percent, y = Region)) +
  geom_point(size = 2.3) +
  geom_errorbarh(aes(xmin = Lower_95_CI, xmax = Upper_95_CI), height = 0.18) +
  scale_x_continuous(limits = c(0, 100), breaks = seq(0, 100, 20), labels = scales::label_percent(scale = 1
  labs(x = "Adjusted EBF probability", y = NULL,
       title = "Adjusted Exclusive Breastfeeding Probability by Region") +
  theme_minimal(base_size = 11)

regional_figure
```

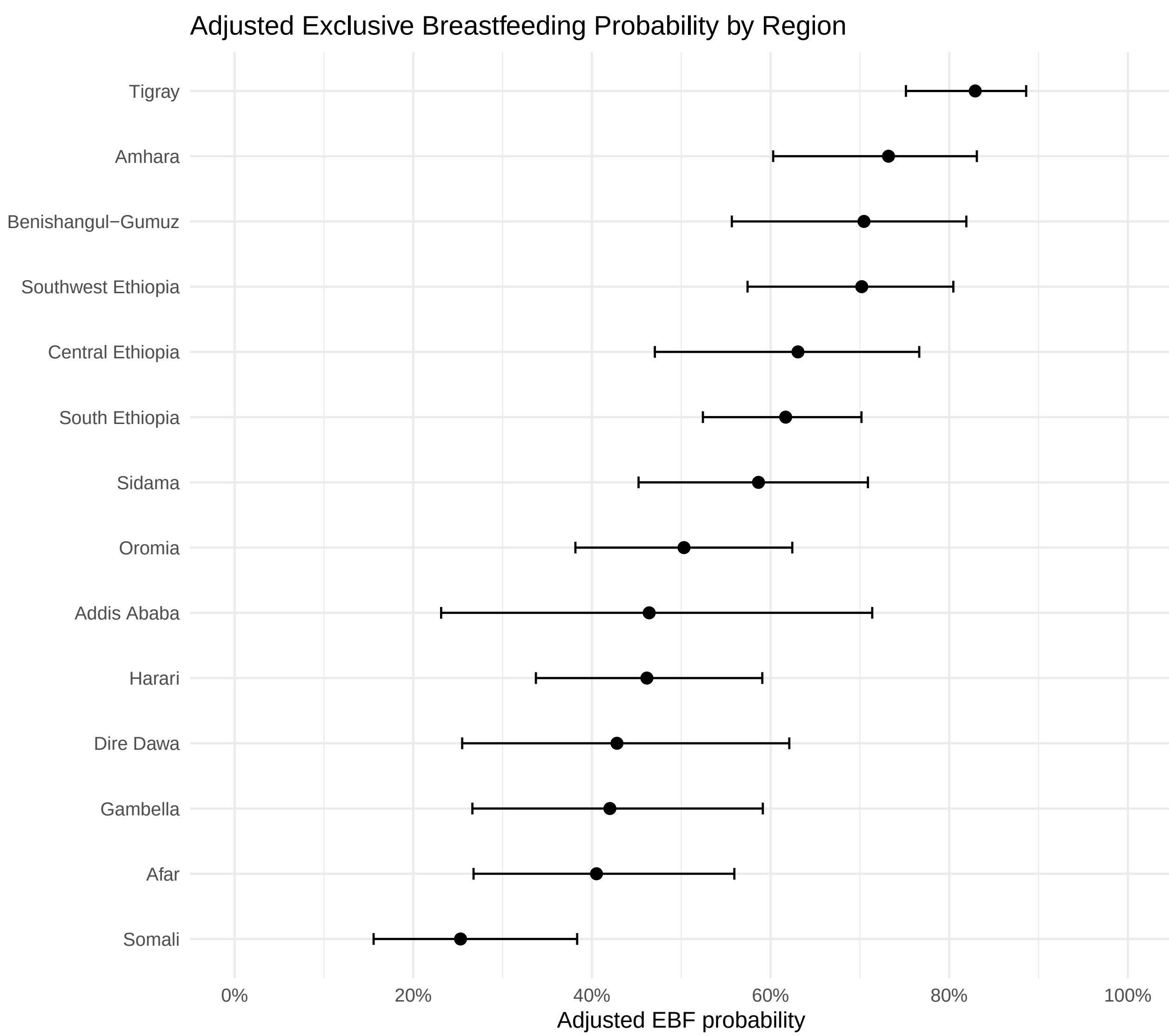

Figure 4: Adjusted predicted probability of exclusive breastfeeding by current region.

```r
ggsave(file.path(output_dir, "figures", "Figure_4_Adjusted_EBF_By_Region.png"), regional_figure,
       width = 8, height = 7, dpi = 300)
```

31.5 Figure 5. Forest plot of adjusted odds ratios

```r
forest_data <- adjusted_or_table %>%
  filter(is.finite(Odds_Ratio), is.finite(Lower_95_CI), is.finite(Upper_95_CI)) %>%
  mutate(Term = forcats::fct_reorder(Term, Odds_Ratio))

forest_figure <- ggplot(forest_data, aes(x = Odds_Ratio, y = Term)) +
  geom_vline(xintercept = 1, linetype = "dashed") +
  geom_point(size = 2) +
  geom_errorbarh(aes(xmin = Lower_95_CI, xmax = Upper_95_CI), height = 0.15) +
  scale_x_log10() +
  labs(x = "Adjusted odds ratio (log scale)", y = NULL,
       title = "Factors Associated with Exclusive Breastfeeding") +
  theme_minimal(base_size = 10)

forest_figure
```

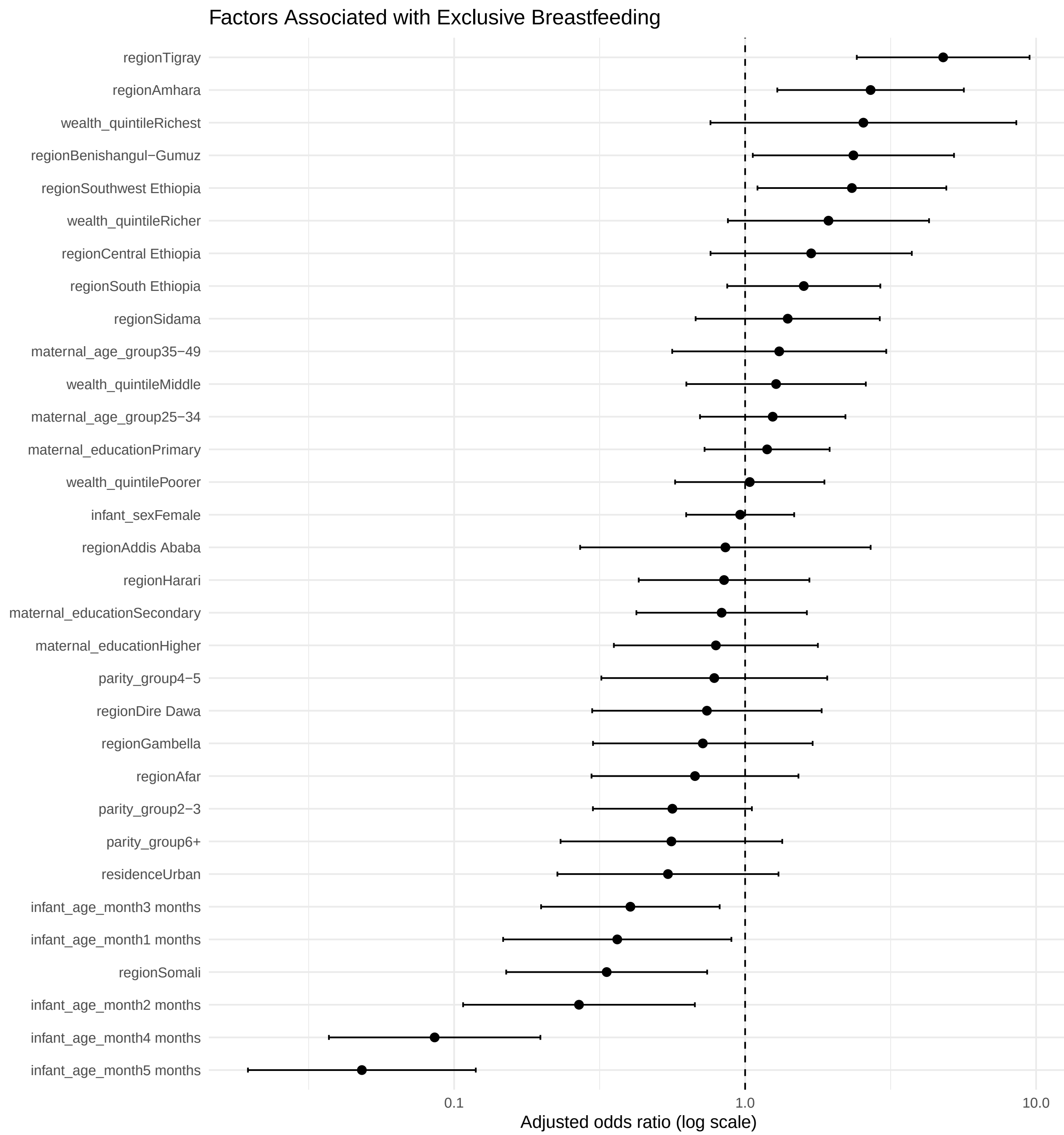

Figure 5: Adjusted odds ratios for factors associated with exclusive breastfeeding.

```
ggsave(file.path(output_dir, "figures", "Figure_5_Adjusted_Odds_Ratio_Forest_Plot.png"), forest_figure,
       width = 8.5, height = 9, dpi = 300)
```

31.6 Final exports and reproducibility record

```
write.csv(region_results, file.path(output_dir, "tables", "Table_Current_Region_Prevalence.csv"), row.names
write.csv(harmonized_region_results, file.path(output_dir, "tables", "Table_Harmonized_Region_Prevalence.cs
write.csv(age_results, file.path(output_dir, "tables", "Table_Age_Specific_Prevalence.csv"), row.names = FA
write.csv(national_results, file.path(output_dir, "tables", "Table_National_Prevalence.csv"), row.names = F
```

```r
write.csv(published_comparison, file.path(output_dir, "tables", "Table_Official_Estimate_Validation.csv"),

saveRDS(
  list(
    analysis_data = analysis_data,
    design = design,
    adjusted_model = adjusted_model,
    national_results = national_results,
    all_prevalence_results = all_prevalence_results,
    adjusted_or_table = adjusted_or_table,
    adjusted_age_probabilities = adjusted_age_probabilities,
    adjusted_region_probabilities = adjusted_region_probabilities
  ),
  file.path(output_dir, "EBF_Analysis_Objects.rds")
)

capture.output(sessionInfo(), file = file.path(output_dir, "R_Session_Info.txt"))

completion_summary <- data.frame(
  Item = c("Analysis completed", "Elapsed minutes", "Output directory", "Final analytical N"),
  Value = c(
    format(Sys.time(), "%Y-%m-%d %H:%M:%S"),
    round(as.numeric(difftime(Sys.time(), session_start, units = "mins")), 2),
    output_dir,
    nrow(analysis_data)
  )
)

report_table(completion_summary, "Analysis completion summary")
```

Table 35: Analysis completion summary

Item	Value
Analysis completed	2026-07-18 19:10:41
Elapsed minutes	0.2
Output directory	C:/Users/aynal/OneDrive/Documents/ETH-DHS-2024-25/EBF_Unified_Output
Final analytical N	1325

32 Interpretation note

The national EBF estimate is a cross-sectional prevalence among infants aged 0-5 completed months based on feeding during the preceding 24 hours. It is not the proportion of a birth cohort that remained continuously exclusively breastfed from birth through six months. Age-specific estimates therefore provide essential information about breastfeeding continuation.