

## Exclusive Breastfeeding — EBF\*

**Abstract:** *Exclusive breastfeeding (EBF) during the first six months of life is a critical determinant of infant health and survival. This study examines EBF patterns in Ethiopia using data from the 2016 Ethiopia Demographic and Health Survey (EDHS), applying a multi-level descriptive framework that integrates national, regional, age-specific, and socioeconomic analyses. The study focuses on living infants aged 0–5 months and uses survey-weighted estimates to ensure national representativeness.*

*Nationally, approximately 58 percent of infants under six months are exclusively breastfed. However, this aggregate figure masks substantial variation across age and region. The most pronounced pattern is a steep age-related decline in EBF, from approximately 75 percent in the first months of life to about 26 percent by month five. This finding indicates that the primary challenge is not initiation of breastfeeding but sustained adherence over time.*

*Regional variation is also substantial, with EBF prevalence ranging from 35.6 percent in Somali to 70.8 percent in Tigray. These differences do not align consistently with urban–rural distinctions or with pastoralist versus non-pastoralist livelihoods, suggesting that EBF is shaped by localized combinations of cultural norms, service access, and maternal practices. Socioeconomic gradients by maternal education and household wealth are present but weak and non-linear, indicating that these factors are not dominant determinants of EBF.*

*The results demonstrate that EBF in Ethiopia is a dynamic, time-sensitive behavior influenced primarily by infant age and context-specific factors rather than by structural socioeconomic conditions. Policy efforts should therefore prioritize supporting the continuation of exclusive breastfeeding beyond the early months of life through targeted, context-specific interventions.*

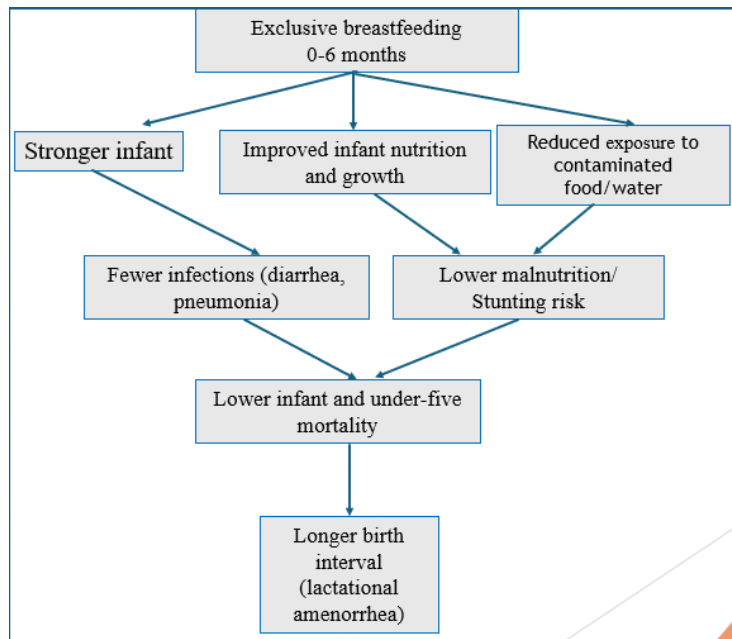
\*By Aynalem Adugna, Research Scientist, State of California

## Introduction

Exclusive breastfeeding (EBF) during the first six months of life is widely recognized as one of the most effective interventions for improving child survival, nutrition, and long-term health outcomes. The World Health Organization (WHO) recommends that infants be exclusively breastfed for the first six months, followed by continued breastfeeding alongside appropriate complementary feeding [1]. EBF has been associated with reductions in infant morbidity and mortality, improved immune protection, and enhanced cognitive development [2-3].

Despite its well-established benefits, EBF remains unevenly practiced across countries and within populations. In many low- and middle-income countries, including Ethiopia, EBF is often initiated but not sustained for the full recommended duration. This pattern reflects the influence of multiple factors, including maternal knowledge, cultural practices, access to health services, employment constraints, and infant feeding norms [4-5].

**Figure 1. Exclusive Breastfeeding Impact Path Pathway**



Unlike women's education (WE), which operates as a long-term structural determinant, EBF is a behavioral outcome that evolves over short time horizons, particularly during the first six months of an infant's life. As such, it is highly sensitive to immediate contextual factors. One of the most consistent empirical findings across settings is the decline in EBF with increasing infant age, reflecting early introduction of complementary foods and liquids [6-7].

In Ethiopia, national estimates suggest that approximately 58 percent of infants under six months are exclusively breastfed, but this aggregate figure masks important variation across regions, age groups, and local contexts [9][24]. Regional disparities are often substantial, and patterns do not always align with conventional socioeconomic gradients such as wealth or maternal education. This suggests that EBF is shaped by a complex interaction of behavioral, cultural, and service-related factors rather than by structural determinants alone.

The central objective is to determine whether EBF exhibits the same type of spatial clustering and structural patterning observed for women's education, or whether it follows a different logic consistent with its behavioral nature. In doing so, the study aims to clarify the extent to which EBF is shaped by place, time (infant age), and socioeconomic context, and to assess the implications for targeted interventions and program design.

## **2. Data and Methods**

### **2.1 Data Source**

This analysis uses data from the Ethiopia Demographic and Health Survey (EDHS) 2016, specifically the Children's Recode (KR) file, which contains information on births in the five years preceding the survey. The KR file is the appropriate dataset for analyzing exclusive

breastfeeding (EBF) because infant feeding practices are recorded at the child level, based on maternal recall of feeding in the 24 hours preceding the survey [8][9].

The EDHS 2016 employs a two-stage stratified sampling design. In the first stage, enumeration areas (clusters) are selected using probability proportional to size. In the second stage, households are systematically sampled within each cluster. All eligible women aged 15–49 in selected households are interviewed, and detailed information is collected for each live birth within the reference period [10][11].

For this analysis, the sample is restricted to:

- Living children aged 0–5 months, consistent with the WHO definition of exclusive breastfeeding
- Children with complete information on feeding practices

This restriction ensures that estimates reflect the population for whom exclusive breastfeeding is recommended and measurable.

## 2.2 Variable Construction

### Exclusive Breastfeeding (EBF)

Exclusive breastfeeding is defined following DHS and WHO guidelines as:

Infants aged 0–5 months who received only breast milk in the 24 hours preceding the survey, with no other liquids or solid foods except for oral rehydration salts, drops, or syrups [8][12].

In DHS data, EBF is constructed using a combination of feeding variables indicating whether the child received:

- water
- other liquids
- formula or milk
- solid or semi-solid foods

### Infant Age (Key Analytical Variable)

Infant age is captured using:

- `b19` = age in months (preferred DHS variable)

This variable is central to the analysis, as EBF is expected to decline with increasing age. Age is treated both as:

- a continuous variable (for interpretation)
- a categorical variable (0, 1, 2, 3, 4, 5 months) for tabulation

## Regional Variable

Region is defined using:

- v101 or labeled factor equivalent (v101\_f)

This variable allows for regional comparisons of EBF prevalence.

## 2.3 Survey Design

As with the women's education analysis, EBF estimates must account for the complex DHS survey design, which includes clustering, stratification, and unequal sampling probabilities.

Survey design was specified using the `survey` package in R:

```
svydesign(  
  id = ~v021,  
  strata = ~v022,  
  weights = ~weight,  
  data = df,  
  nest = TRUE  
)
```

Where:

- v021 = cluster (primary sampling unit)
- v022 = sampling strata
- v005 = sampling weight (normalized by dividing by 1,000,000)

Survey-weighted estimation was used for:

- national EBF prevalence
- regional comparisons
- age-specific EBF estimates

This ensures that results are representative of the national population of infants aged 0–5 months [10][11].

## 3. Results

### 3.1 National Distribution

The national prevalence of exclusive breastfeeding (EBF) among infants aged 0–5 months in Ethiopia was estimated at approximately 58 percent, consistent with the 2016 Ethiopia Demographic and Health Survey (EDHS) report and with prior DHS-based analyses [4][6]. This estimate confirms that, at the national level, a majority of infants under six months are still exclusively breastfed, but a substantial minority are not.

This national figure is important as a benchmark, but on its own it can be misleading. As with women’s education, national averages conceal important forms of heterogeneity. In the case of EBF, the hidden variation is less about stable long-run structural disparities and more about differences across regions, infant age, and immediate maternal and household circumstances. The national estimate, therefore, serves as a useful starting point, but not as an adequate summary of the Ethiopian breastfeeding environment.

### 3.2 Regional Distribution

Regional analysis reveals substantial variation in EBF prevalence across Ethiopia. Weighted estimates show the following regional pattern:

**Table 1. Exclusive Breastfeeding Regional Prevalence**

Region	EBF (%)
Tigray	70.8
Afar	53.1
Amhara	63.8
Oromia	56.9
Somali	35.6
Benishangul-Gumuz	66.5
SNNPR	55.8
Gambela	51.3
Harari	40.7
Addis Ababa	56.0
Dire Dawa	57.4

Several features of this distribution deserve emphasis. Substantively, the national prevalence indicates that Ethiopia performs neither exceptionally poorly nor exceptionally well in absolute terms. Rather, it occupies an intermediate position in which EBF is widely initiated but insufficiently sustained. This interpretation is reinforced by the age-pattern analysis presented below, which shows that the apparent national success is driven largely by very high EBF prevalence in the earliest months of life, followed by sharp declines thereafter.

First, the range is wide, extending from 35.6 percent in Somali to 70.8 percent in Tigray. This confirms that EBF is not evenly distributed across the country and that regional context matters. However, the pattern is not reducible to a simple urban–rural distinction. Addis Ababa, despite being the capital and the most urbanized region, does not rank highest; instead, its EBF prevalence is close to the national average. Similarly, Harari, which is predominantly urban, exhibits a comparatively low EBF level, closer to that of Somali than to that of Addis Ababa. This indicates that urbanization alone does not explain the pattern.

Second, pastoralist or highly mobile settings also do not yield a single consistent EBF profile. Afar, despite its strong mobility and pastoralist character, has a moderate EBF prevalence (53.1 percent), notably higher than Somali. This shows that similar livelihood systems can yield

different breastfeeding outcomes depending on their interactions with health services, local norms, maternal support systems, and infant feeding beliefs.

Third, several regions cluster around the national mean, including Oromia, SNNPR, Addis Ababa, and Dire Dawa. This suggests that broad national averages are partly sustained by a large middle range of regions rather than by uniformity.

Taken together, the regional results suggest that EBF in Ethiopia is shaped by context-specific combinations of factors, rather than by one dominant geographic or socioeconomic axis. This interpretation is consistent with prior work showing substantial regional heterogeneity in breastfeeding practices across Ethiopia [6][7].

### 3.3 Age-Based Decline

The most analytically important finding in the EBF analysis is the sharp decline in EBF with infant age. Weighted estimates by completed month of age show the following pattern:

**Table 2. Exclusive Breastfeeding Percentage by Infant Age**

Age (months)	EBF (%)
0	74.4
1	74.8
2	60.1
3	68.2
4	46.7
5	26.0

This pattern shows that EBF is high in the first two months of life, remains relatively elevated at month 3, and then falls sharply by months 4 and 5. The decline from approximately 75 percent at 0–1 months to 26 percent at 5 months is striking and provides the clearest explanation for why the national EBF estimate can appear moderately strong while sustained six-month exclusivity remains weak.

Substantively, this confirms that the principal EBF challenge in Ethiopia is continuation rather than initiation. Mothers overwhelmingly begin breastfeeding, and many maintain exclusivity in the earliest months, but continuation becomes increasingly difficult as infants age. This pattern is consistent with the broader literature on exclusive breastfeeding in Ethiopia and comparable low- and middle-income settings, where the introduction of water, other liquids, or complementary foods typically increases as infants grow older [2][5][6].

The slight increase at month 3 relative to month 2 should not be over-interpreted. It may reflect sampling variability rather than a substantive behavioral reversal. The dominant pattern is unequivocal: EBF declines markedly over the first six months of life.

From a policy perspective, this finding is especially important because it shifts attention away from broad awareness messaging toward targeted support for sustaining EBF between months 2 and 5, when drop-off becomes most pronounced.

### 3.4 Socioeconomic Gradients

#### Maternal Education

Maternal education does not show a simple, linear relationship with EBF. Weighted estimates indicate the following pattern:

**Table 3. Exclusive Breastfeeding Percentage by Mothers' Educational Status**

Education	EBF (%)
No education	57.8
Primary	60.1
Secondary	55.8
Higher	41.7

This pattern is clearly non-linear. EBF prevalence is slightly higher among mothers with primary education than among those with no education, but it declines again at the secondary level and drops markedly among mothers with higher education. This suggests that education does not operate as a straightforward positive predictor of EBF.

The likely explanation is that education works through competing pathways. On the one hand, education can improve health knowledge, service utilization, and understanding of recommended infant feeding practices. On the other hand, higher levels of education are often associated with employment, urban residence, time scarcity, and earlier return to work, all of which may reduce the practical feasibility of exclusive breastfeeding. In this sense, education is better understood as a context-setting variable than as a direct determinant of EBF.

#### Wealth

The wealth gradient is also relatively modest:

**Table 4. Exclusive Breastfeeding Percentage by Wealth Quartile**

Wealth Quintile	EBF (%)
Poorest	~60
Poorer	~59
Middle	~58
Richer	~55
Richest	~50

The overall pattern is one of gradual decline with increasing wealth, but the magnitude is limited. The difference between the poorest and richest groups is real, but not large enough to suggest

that wealth is a dominant driver of EBF. This aligns with the broader interpretation that wealth operates through offsetting mechanisms: greater wealth may improve access to care and information, but it may also increase access to breastmilk substitutes and heighten the opportunity costs of continued exclusive breastfeeding [6].

Together, the education and wealth findings suggest that EBF in Ethiopia is not primarily structured by socioeconomic gradients in a conventional sense. Socioeconomic position matters, but it does not explain the strongest patterns observed in the data. Those stronger patterns are age-related decline and context-specific regional variation.

## **4. Discussion**

### **4.1 EBF as a Behavioral System**

The results of this analysis reinforce the interpretation of exclusive breastfeeding (EBF) as a behavioral system shaped by short-term, context-specific factors, rather than as a structural outcome determined by long-run socioeconomic conditions. This distinction is critical. Unlike women's education—which reflects cumulative investments in schooling infrastructure, institutional capacity, and intergenerational transmission—EBF operates within a narrow temporal window and is highly sensitive to immediate maternal circumstances, infant age, and feeding practices [13][14].

The empirical patterns observed in Ethiopia align with this conceptualization. National prevalence appears moderately strong at approximately 58 percent, but this aggregate masks rapid behavioral change over the first six months of life. The defining feature of EBF in Ethiopia is not whether breastfeeding is initiated—initiation is nearly universal—but whether exclusivity is sustained. This distinction shifts analytical emphasis from structural access to continuity of practice, which is inherently more fragile and more dependent on daily decision-making.

This behavioral framing helps explain why EBF does not map cleanly onto conventional indicators such as wealth and education. These variables may influence knowledge and access to services, but they do not fully capture the constraints and trade-offs that mothers face in sustaining exclusive breastfeeding over time.

### **4.2 Age as the Dominant Determinant**

Among all variables examined, infant age emerges as the most powerful and consistent determinant of EBF. The observed decline from approximately 75 percent in the first months of life to 26 percent by month five is both large in magnitude and consistent with findings from DHS-based studies in Ethiopia and comparable settings [13][14].

This age gradient reflects a transition from early postnatal conditions—when breastfeeding is biologically and socially supported—to later months, when competing pressures intensify. These pressures may include:

- perceptions of insufficient breast milk as the infant grows
- introduction of water or other liquids due to cultural norms
- increasing maternal workload or return to employment
- advice from family members or community norms regarding complementary feeding

Importantly, the age effect is not merely a statistical pattern; it is a behavioral trajectory. EBF is initially the default state but becomes progressively harder to maintain. This suggests that interventions should focus less on initiation—which is already high—and more on preventing early discontinuation, particularly after the second or third month of life.

### **4.3 Regional Variation Without Simple Geographic Logic**

The regional analysis demonstrates that EBF varies substantially across Ethiopia, but the pattern does not conform to a single explanatory framework. Regions such as Tigray and Benishangul-Gumuz exhibit relatively high EBF prevalence, while Somali and Harari show much lower levels. However, these differences cannot be explained solely by urbanization, wealth, or livelihood systems.

For example:

- Addis Ababa, despite its urban and service-rich environment, exhibits EBF levels close to the national average rather than at the top of the distribution
- Harari, also predominantly urban, shows relatively low EBF
- Afar, a largely pastoralist and mobile region, exhibits moderate EBF levels that exceed those observed in Somali

These patterns indicate that EBF is shaped by localized combinations of cultural norms, service access, maternal support systems, and feeding practices, rather than by broad geographic categories. The absence of a simple urban–rural or pastoralist–non-pastoralist gradient suggests that breastfeeding behavior is highly context-dependent.

This interpretation is consistent with prior studies that emphasize the role of community-level practices and norms in shaping infant feeding behavior, often overriding socioeconomic gradients [6][7].

### **4.4 Weak and Non-Linear Socioeconomic Gradients**

The analysis finds that both education and wealth are associated with EBF, but their effects are weak, non-linear, and sometimes counterintuitive.

Maternal education shows a pattern in which EBF increases slightly from no education to primary education, then declines at higher levels of education. Similarly, wealth is associated with a gradual decline in EBF from the poorest to the richest households. These patterns suggest that socioeconomic status does not operate as a straightforward facilitator of exclusive breastfeeding.

Instead, education and wealth appear to influence EBF through offsetting mechanisms:

- Positive pathways: improved knowledge, greater exposure to health messaging, increased access to services
- Negative pathways: employment constraints, time scarcity, increased access to breastmilk substitutes, changing norms in more urbanized or affluent settings

The net effect is a flattened or inverted gradient, in which higher socioeconomic status does not consistently translate into better EBF outcomes. This finding directly challenges interpretations that treat education and wealth as primary determinants of breastfeeding behavior.

## **4.5 Implications for Interpreting Regression-Based Studies**

These findings have important implications for the interpretation of regression-based analyses of EBF, including those that report statistically significant associations between EBF and variables such as education, wealth, media exposure, or region.

First, many of these variables are highly correlated. Education, wealth, and media access are not independent dimensions; they often move together as part of a broader socioeconomic profile. When included simultaneously in regression models without careful assessment of collinearity, coefficient estimates may be unstable or misleading.

Second, regression models typically impose a global structure, assuming that relationships between variables are uniform across the entire population. The results presented here suggest that this assumption may not hold for EBF. The importance of age, the variability across regions, and the weak socioeconomic gradients all point to a system in which relationships are context-specific and may differ across subpopulations.

Third, regression coefficients are often interpreted as evidence of causal influence, even when the underlying relationships are complex and mediated by unobserved factors. For example, the association between higher education and lower EBF does not imply that education reduces breastfeeding; rather, it reflects the broader set of conditions associated with higher education, including employment and lifestyle factors.

Taken together, these considerations suggest that EBF is better understood through a combination of descriptive, stratified, and context-sensitive analyses rather than relying solely on multivariable regression.

## **4.6 Policy Implications**

The findings of this study point to several clear policy implications.

First, interventions should prioritize sustaining EBF beyond the early months of life, rather than focusing primarily on initiation. Given the steep decline after months two and three, targeted support during this period is likely to yield the greatest gains.

Second, regional variation suggests the need for context-specific strategies. Interventions that are effective in one region may not translate directly to another, given differences in cultural norms, service access, and maternal support systems.

Third, the weak socioeconomic gradients indicate that EBF should not be treated solely as a function of poverty or education. Instead, policies should address practical constraints and behavioral drivers, such as:

- workplace policies that support breastfeeding
- counseling on perceived milk insufficiency
- community-based support systems
- norms around early introduction of liquids and foods

Finally, the absence of sufficient cluster-level data for spatial analysis highlights the importance of data-aware program design. Monitoring systems and future surveys could be strengthened by ensuring adequate representation of infants in the 0–5-month age range to support more granular analysis.

## 5. Conclusion

This study examined exclusive breastfeeding (EBF) in Ethiopia using a multi-level descriptive framework, drawing on nationally representative data from the EDHS 2016. The analysis focused on national prevalence, regional variation, age-specific patterns, and socioeconomic gradients, while explicitly recognizing the limits imposed by the age-restricted sample on cluster-level spatial inference. The results provide a clear and internally consistent picture of EBF as a behavioral system characterized by strong temporal dynamics and context-specific variation.

At the national level, EBF prevalence of approximately 58 percent indicates that breastfeeding is widely practiced and that exclusivity is achieved for a substantial proportion of infants. However, this aggregate measure masks the most important underlying pattern: a pronounced decline in EBF with increasing infant age. The drop from roughly three-quarters of infants being exclusively breastfed in the first months of life to approximately one-quarter by month five reveals that the central challenge is not initiation but sustained adherence to exclusive breastfeeding over time.

Regional variation further demonstrates that EBF is unevenly distributed across Ethiopia, but not in a way that conforms to simple geographic or socioeconomic classifications. Differences between regions such as Somali, Harari, Afar, and Addis Ababa cannot be explained solely by urbanization, livelihood systems, or access to services. Instead, they reflect localized combinations of cultural norms, maternal practices, service environments, and social support structures. This reinforces the conclusion that EBF is inherently context-dependent.

Socioeconomic gradients, particularly those related to maternal education and household wealth, are present but weak and non-linear. Higher levels of education and wealth do not consistently correspond to higher EBF prevalence. Rather, they appear to operate through competing

mechanisms that both enable and constrain breastfeeding. These findings challenge interpretations that treat education and wealth as dominant or straightforward determinants of EBF, underscoring the need for more nuanced analytical approaches.

Methodologically, the study demonstrates the value of aligning analytical techniques with the data structure. The decision not to pursue cluster-level hot-spot analysis—due to the limited number of clusters with infants aged 0–5 months—ensures that conclusions are grounded in reliable evidence. This highlights a broader principle: robust inference depends not only on analytical sophistication but also on appropriate matching between methods and data.

From a policy perspective, the findings underscore the importance of shifting emphasis from awareness and initiation to supporting the continuation of exclusive breastfeeding, particularly after the first two months of life. Interventions should be tailored to regional contexts and should address the practical and behavioral constraints that mothers face, including workload, perceptions of milk sufficiency, and social norms regarding infant feeding.

In conclusion, exclusive breastfeeding in Ethiopia is best understood not as a static indicator or a simple function of socioeconomic status, but as a dynamic, time-sensitive behavior shaped by interacting individual, cultural, and contextual factors. Effective policy and program responses must therefore move beyond one-size-fits-all approaches and instead engage with the specific conditions that influence breastfeeding practices across different settings and stages of early infancy.

Nationally, approximately 58 percent of infants under six months are exclusively breastfed. However, this aggregate figure masks substantial variation across age and region. The most pronounced pattern is a steep age-related decline in EBF, from approximately 75 percent in the first months of life to about 26 percent by month five. This finding indicates that the primary challenge is not initiation of breastfeeding but sustained adherence over time.

Regional variation is also substantial, with EBF prevalence ranging from 35.6 percent in Somali to 70.8 percent in Tigray. These differences do not align consistently with urban–rural distinctions or with pastoralist versus non-pastoralist livelihoods, suggesting that EBF is shaped by localized combinations of cultural norms, service access, and maternal practices. Socioeconomic gradients by maternal education and household wealth are present but weak and non-linear, indicating that these factors are not dominant determinants of EBF.

The results demonstrate that EBF in Ethiopia is a dynamic, time-sensitive behavior influenced primarily by infant age and context-specific factors rather than by structural socioeconomic conditions. Policy efforts should therefore prioritize supporting the continuation of exclusive breastfeeding beyond the early months of life through targeted, context-specific interventions.

# Appendix 1: Methodological Critique of EDHS Regression Findings - Example Muluneh (2023)

Mitiku Wale Muluneh. Determinants of exclusive breastfeeding practices among mothers in Ethiopia. Published: February 9, 2023. <https://doi.org/10.1371/journal.pone.0281576>

“Results: The prevalence of EBF was 58% of infants under age 6 months. Mothers age 25–34 (AOR = 1.74; 95% CI 1.31–2.32), child age 4–5 months (AOR = 0.74; 95% CI 0.66–0.84), married marital status (AOR = 1.26; 95% CI 1.06–1.50), mothers attained secondary education or higher (AOR = 2.00; 95% CI 1.54–2.58), husband attained secondary education or higher (AOR = 1.70; 1.39–2.13), richer wealth index (AOR = 0.35; 95% CI 0.18–0.69), accessed to the media (AOR = 1.77; 95% CI 1.38–2.27), number of living children 3–4 (AOR = 0.49; 95% CI 0.25–0.95), health facility (AOR = 1.87; 95% CI 1.09–3.20), rural residence (AOR = 0.66; 95% CI 0.49–0.89) and mothers living in Afar (AOR = 100.2; 95% CI 15.68–640.61), Somali (AOR = 52.65; 95% CI 8.48–326.77), SNNPR (AOR = 6.94; 95% CI 1.05–45.79), Harari (AOR = 61.94; 95% CI 9.75–393.44), Addis Ababa (AOR = 13.07; 95% CI 2.06–82.99), and Dire Dawa (AOR = 28.91; 95% CI 4.38–190.86) were associated with EBF practice.”

Muluneh (2023) reports a series of adjusted odds ratios (AORs) for exclusive breastfeeding (EBF), including effects for maternal age, education, wealth, media exposure, residence, and region. While the reported national prevalence (58 percent) aligns closely with the descriptive findings presented in this study, several aspects of the regression results raise important methodological concerns.

## 1. Use of “Determinants” and Implied Causality

The study interprets statistically significant associations as *determinants* of EBF. However, the underlying data are cross-sectional DHS data, which do not support causal inference. The reported AORs reflect conditional associations given the model specification, not causal effects.

For example, findings such as:

- maternal secondary education or higher (AOR = 2.00)
- media exposure (AOR = 1.77)
- health facility delivery (AOR = 1.87)

are presented as if they independently drive EBF behavior. In reality, these variables are part of broader socioeconomic and contextual systems. Without temporal ordering or causal identification, the interpretation should be limited to associations, not determinants.

## 2. Multicollinearity and Non-Independence of Covariates

A central concern is the inclusion of multiple highly correlated variables in the same model without reported diagnostics for multicollinearity.

Key examples include:

- Maternal education, husband's education, and wealth index
- Wealth, media exposure, and urban/rural residence
- Region and residence, which jointly capture geographic and socioeconomic variation

These variables are not independent dimensions; they reflect overlapping aspects of socioeconomic status and access to services. When included simultaneously in a logistic regression model:

- coefficients represent partial effects conditional on other correlated variables
- estimates can become unstable or attenuated
- interpretation as independent effects becomes problematic

The absence of reported diagnostics (e.g., variance inflation factors or correlation matrices) makes it difficult to assess whether the estimated AORs are statistically reliable.

### **3. Implausibly Large Regional Effects**

The most striking results are the extremely large regional odds ratios, including:

- Afar: AOR = 100.2 (95% CI: 15.68–640.61)
- Somali: AOR = 52.65 (95% CI: 8.48–326.77)
- Harari: AOR = 61.94 (95% CI: 9.75–393.44)
- Dire Dawa: AOR = 28.91 (95% CI: 4.38–190.86)
- Addis Ababa: AOR = 13.07 (95% CI: 2.06–82.99)

These magnitudes are substantively implausible for a behavioral outcome such as EBF. Odds ratios of this scale suggest that model estimates are being driven by:

- small sample sizes in certain regions
- sparse data within categories
- quasi-separation or complete separation in the logistic model
- or model overfitting with too many correlated predictors

The extremely wide confidence intervals further indicate low precision and instability. Rather than reflecting true effect sizes, these values likely signal model specification problems.

### **4. Inconsistency with Descriptive Patterns**

The regression results also diverge from well-established descriptive patterns observed in DHS data and confirmed in this study.

## **Age effect**

Muluneh reports:

- child age 4–5 months (AOR = 0.74)

While this correctly identifies a negative association, it substantially understates the magnitude of decline observed descriptively. In the present analysis, EBF falls from approximately 75 percent in early months to 26 percent by month five—a much sharper gradient than suggested by the odds ratio.

## **Education and wealth**

Muluneh reports:

- maternal education (AOR = 2.00, positive)
- richer wealth (AOR = 0.35, negative)

These results imply strong and opposing effects. However, the descriptive analysis shows:

- non-linear education patterns, with lower EBF among the highest education group
- modest wealth gradients, not large enough to support strong causal claims

This inconsistency suggests that the regression model may be misallocating variance across correlated predictors, rather than capturing stable underlying relationships.

## **5. Omitted Contextual and Behavioral Factors**

The model includes standard DHS covariates but omits key determinants of breastfeeding behavior, such as:

- cultural norms regarding infant feeding
- perceptions of breast milk sufficiency
- maternal workload and time constraints
- social and family support systems

Because these factors are not directly measured, their effects may be absorbed into observed variables such as education, wealth, or region. This introduces omitted variable bias, further complicating interpretation.

## **6. Interpretation of Model Outputs**

Taken together, the reported AORs should be interpreted with caution:

- They represent associations conditional on a potentially misspecified model
- Large coefficients, particularly for the region, likely reflect data and modeling artifacts

- The apparent importance of certain variables may be overstated due to collinearity and omitted context

Rather than identifying independent determinants, the model appears to capture a complex, overlapping system of correlated influences, which is not well represented by standard logistic regression coefficients.

## **7. Implications for EBF Analysis**

The descriptive findings presented in this study offer a more coherent interpretation of EBF patterns in Ethiopia:

- Infant age is the dominant and most consistent factor
- Regional variation exists, but does not align with simple categories
- Education and wealth effects are weak and non-linear

These patterns suggest that EBF is best understood as a dynamic behavioral process, not as an outcome determined by a small set of independent predictors.

## **8. Concluding Assessment**

Muluneh (2023) illustrates common challenges in applying regression models to DHS data. The combination of causal language, untested multicollinearity, and implausibly large coefficients limits the interpretability of the reported findings.

A more appropriate analytical approach emphasizes:

- careful distinction between association and causation
- explicit assessment of variable interdependence
- reliance on descriptive and stratified analyses to capture context-specific patterns

Such an approach yields results that are both statistically defensible and substantively meaningful, particularly for complex behaviors such as exclusive breastfeeding.

## Appendix 2: Core R Workflow for Exclusive Breastfeeding (Teaching Version with Line-by-Line Notes)

### A. What this script does

This script does six main things:

1. Opens the Ethiopia DHS 2016 Birth Recode file
2. Keeps only the variables needed for exclusive breastfeeding (EBF) analysis
3. Restricts the data to living infants aged 0–5 months
4. Builds an EBF variable from breastfeeding and feeding-practice variables
5. Creates survey-weighted national and subgroup estimates
6. Checks cluster counts to determine whether spatial analysis is feasible

### B. Very important symbols before you begin

#### 1. What does `<-` mean?

Example:

```
path <- "C:/Data/ETBR71FL.DTA"
```

Read this as:

“Store this value in the object called `path`”

So `<-` means:

- “put into”
- “save as”
- “assign to”

#### 2. What does `%>%` mean?

Example:

```
data %>%  
  filter(x == 1) %>%  
  select(y, z)
```

Read `%>%` as:

- “then”
- “and next do this”
- “take the result from the previous step and pass it to the next step”

So the example means:

1. start with `data`
2. then keep only rows where `x == 1`
3. then keep only columns `y` and `z`

This helps R code read like a recipe.

### 3. What does `$` mean?

Example:

```
df$ebf
```

This means:

the variable `ebf` inside the dataset `df`

So `$` means:

- “inside”
- “from this dataset, get this variable”

### 4. What does `~` mean?

Example:

```
id = ~v021
```

This is called a formula. In survey analysis, it means:

use the variable `v021`

So `~` can be read as:

- “use variable”
- “model this using”

## C. Load the packages

```
# Load haven.  
# haven lets R read Stata files such as DHS .DTA files.  
library(haven)  
  
# Load dplyr.  
# dplyr is used for selecting variables, creating variables,  
# filtering rows, grouping, and summarising.
```

```

library(dplyr)

# Load survey.
# survey is needed because DHS data use a complex sample design.
# Without this package, weighted percentages would be incorrect.
library(survey)

# Load forcats.
# forcats helps manage categorical variables (factors).
library(forcats)

```

## D. Read the DHS Birth Recode file

```

# Create an object called path.
# This stores the address of the DHS Birth Recode file.
path <-
"C:/Users/aynal/OneDrive/Documents/DHS2016/Data/ETBR71DT/ETBR71FL.DTA"

# Read the Stata file into R.
# Store the result in an object called br.
# "br" stands for Birth Recode.
br <- read_dta(path)

```

### How to read this

- `path <- ...` means save the file location as `path`
- `br <- read_dta(path)` means open the file and store it as `br`

## E. Select only the variables needed

```

# Create a smaller working dataset called ebf_vars.
# This keeps only the variables needed for EBF analysis.
ebf_vars <- br %>%
  select(
    caseid,                # case identifier

    v001, v021, v022,      # cluster ID and survey design variables
    v024,                  # region
    v005,                  # DHS sampling weight

    b5,                   # child alive
    b19,                  # child age in months
    b4,                   # child's sex

    v404,                 # currently breastfeeding
    v409, v409a,          # water / water-based liquids
    v410, v410a,          # other liquids
    v411, v411a,          # milk / formula
    m38,                  # bottle feeding
    m39, m39a,            # foods / feeding frequency

    m34,                  # early initiation of breastfeeding
  )

```

```

v106,          # mother's education
v190,          # household wealth
v025,          # urban/rural residence
v012,          # mother's age
v201           # parity (children ever born)
)

```

## Why this step matters

DHS files contain hundreds or thousands of variables. Keeping only the needed variables:

- makes the dataset easier to work with
- reduces confusion
- improves speed

## F. Restrict to the EBF analytic population and build the EBF variable

```

# Create the main analysis dataset called df.
df <- ebf_vars %>%

# Keep only living children.
# b5 == 1 means the child is alive.
filter(b5 == 1) %>%

# Keep only infants 0-5 months old.
# This matches the WHO definition of the age group for EBF.
filter(b19 < 6) %>%

# Create new variables using mutate().
mutate(

# Convert DHS weight to usable form.
# DHS weights must be divided by 1,000,000.
weight = v005 / 1000000,

# bf = currently breastfeeding
# v404 == 1 means yes
bf = ifelse(v404 == 1, 1, 0),

# Replace missing values with 0 before creating feeding indicators.
# This is important because missing values can otherwise create NA
errors.
v409 = ifelse(is.na(v409), 0, v409),
v409a = ifelse(is.na(v409a), 0, v409a),
v410 = ifelse(is.na(v410), 0, v410),
v410a = ifelse(is.na(v410a), 0, v410a),
v411 = ifelse(is.na(v411), 0, v411),
v411a = ifelse(is.na(v411a), 0, v411a),
m38 = ifelse(is.na(m38), 0, m38),
m39 = ifelse(is.na(m39), 0, m39),
m39a = ifelse(is.na(m39a), 0, m39a),

# Create simple yes/no indicators for non-breastmilk intake.
water = ifelse(v409 == 1 | v409a == 1, 1, 0),

```

```

liquids = ifelse(v410 == 1 | v410a == 1, 1, 0),
milk     = ifelse(v411 == 1 | v411a == 1, 1, 0),
bottle   = ifelse(m38 == 1, 1, 0),

# foods = any solid / semi-solid feeding
# m39a == 1 means solid or semi-solid food yesterday
# m39 > 0 can indicate any non-breastmilk feeding frequency
foods = ifelse(m39a == 1 | m39 > 0, 1, 0),

# any_non_bf = whether the infant had anything besides breastmilk
any_non_bf = ifelse(
  water == 1 | liquids == 1 | milk == 1 | bottle == 1 | foods == 1,
  1, 0
),

# Final EBF variable.
# ebf = 1 only if:
# - infant is currently breastfeeding
# - infant had no other liquids or foods
ebf = case_when(
  bf == 1 & any_non_bf == 0 ~ 1,
  bf == 1 & any_non_bf == 1 ~ 0,
  bf == 0 ~ 0,
  TRUE ~ NA_real_
)
)

```

## How to read `ifelse()`

Example:

```
bf = ifelse(v404 == 1, 1, 0)
```

This means:

- if `v404 == 1`, then assign 1
- otherwise assign 0

Read it as:

“If this condition is true, use the first value; if not, use the second”

## How to read `case_when()`

Example:

```

ebf = case_when(
  bf == 1 & any_non_bf == 0 ~ 1,
  bf == 1 & any_non_bf == 1 ~ 0,
  bf == 0 ~ 0,
  TRUE ~ NA_real_
)

```

This is like a series of “if-then” rules.

Read the first line as:

If the infant is breastfeeding and had no other foods or liquids, then EBF = 1

Read the second line as:

If the infant is breastfeeding but also had other foods or liquids, then EBF = 0

## G. Check whether the EBF variable worked

```
# Count how many infants are coded as EBF = 1 or EBF = 0
table(df$ebf, useNA = "ifany")

# Show proportions, excluding missing values
prop.table(table(df$ebf, useNA = "no"))
```

## Why this is important

Before doing any analysis, always check:

- Did the variable get created correctly?
- Are both 0 and 1 present?
- Are there too many missing values?

## H. Create the survey design object

```
# Build the DHS survey design object.
design <- svydesign(
  id = ~v021,          # cluster / primary sampling unit
  strata = ~v022,     # sampling strata
  weights = ~weight,  # DHS normalized weights
  data = df,          # the EBF dataset
  nest = TRUE
)
```

## Why this step matters

DHS is not a simple random sample. It is:

- clustered
- stratified
- weighted

So R must be told how the sample was designed.

Without this step, your weighted estimates will not be correct.

## I. National EBF prevalence

```
# Calculate the national weighted EBF prevalence.
svymean(~ebf, design, na.rm = TRUE)
```

### How to read this

- `svymean()` = survey-weighted mean
- `~ebf` = use the variable `ebf`
- `design` = use the DHS survey design
- `na.rm = TRUE` = remove missing values

Because `ebf` is coded 0/1:

- `mean(ebf)` = proportion exclusively breastfed

## J. Regional EBF distribution

```
# Convert region codes into readable labels
df <- df %>%
  mutate(region = as_factor(v024))

# Rebuild survey design after adding region factor
design <- svydesign(
  id = ~v021,
  strata = ~v022,
  weights = ~weight,
  data = df,
  nest = TRUE
)

# Calculate weighted EBF by region
ebf_region <- svyby(
  ~ebf,
  ~region,
  design,
  svymean,
  na.rm = TRUE,
  vartype = c("se", "ci")
)

# Print the table
print(ebf_region)
```

### How to read `svyby()`

Read it as:

“Calculate survey-weighted EBF, by region”

## K. Age-based EBF decline

```
# Calculate weighted EBF by infant age in completed months
ebf_age <- svyby(
  ~ebf,
  ~b19,
  design,
  svymean,
  na.rm = TRUE,
  vartype = c("se", "ci")
)

# Print the age table
print(ebf_age)
```

## Why this step matters

This is one of the most important analyses in the chapter.

It shows:

- how EBF changes from month 0 to month 5
- whether the main issue is initiation or continuation

## L. Maternal education and EBF

```
# Create readable education labels
df <- df %>%
  mutate(
    v106_f = factor(
      v106,
      levels = c(0, 1, 2, 3),
      labels = c("No education", "Primary", "Secondary", "Higher")
    )
  )

# Rebuild survey design after adding education factor
design <- svydesign(
  id = ~v021,
  strata = ~v022,
  weights = ~weight,
  data = df,
  nest = TRUE
)

# Weighted EBF by maternal education
ebf_edu <- svyby(
  ~ebf,
  ~v106_f,
  design,
  svymean,
  na.rm = TRUE,
  vartype = c("se", "ci")
)
```

```
)  
print(ebf_edu)
```

## Why do this?

This shows whether EBF:

- rises with education
- falls with education
- or is non-linear

## M. Wealth gradient

```
# Create readable wealth labels  
df <- df %>%  
  mutate(  
    v190_f = factor(  
      v190,  
      levels = 1:5,  
      labels = c("Poorest", "Poorer", "Middle", "Richer", "Richest")  
    )  
  )  
  
# Rebuild survey design after adding wealth factor  
design <- svydesign(  
  id = ~v021,  
  strata = ~v022,  
  weights = ~weight,  
  data = df,  
  nest = TRUE  
)  
  
# Weighted EBF by wealth quintile  
ebf_wealth <- svyby(  
  ~ebf,  
  ~v190_f,  
  design,  
  svymean,  
  na.rm = TRUE,  
  vartype = c("se", "ci")  
)  
  
print(ebf_wealth)
```

## N. Why cluster-level spatial analysis was not pursued

```
# Count the number of unique clusters represented  
n_distinct(df$v001)  
  
# Count infants by cluster  
cluster_ebf_counts <- df %>%  
  group_by(v001) %>%
```

```
summarise(  
  n = n(),  
  .groups = "drop"  
)  
  
# Number of clusters with eligible infants  
nrow(cluster_ebf_counts)  
  
# Look at the first few clusters  
head(cluster_ebf_counts)
```

## Why this matters

This step tells you whether there are enough clusters for spatial analysis.

In your case, once the sample was restricted to infants aged 0–5 months:

- only about **46 clusters** had eligible infants

That is too sparse for reliable national hot-spot analysis.

## References

- [1] World Health Organization (WHO). (n.d.). *Infant and young child feeding*. Available at: <https://www.who.int/health-topics/infant-and-young-child-feeding>
- [2] Victora, C. G., et al. (2016). Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effect. *The Lancet*, 387(10017), 475–490. Available at: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(15\)01024-7/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)01024-7/fulltext)
- [3] UNICEF. (n.d.). *Breastfeeding: A mother's gift for every child*. Available at: [https://www.unicef.org/nutrition/index\\_24824.html](https://www.unicef.org/nutrition/index_24824.html)
- [4] Central Statistical Agency (CSA) [Ethiopia] and ICF. (2016). *Ethiopia Demographic and Health Survey 2016: Final Report*. Available at: <https://dhsprogram.com/publications/publication-FR328-DHS-Final-Reports.cfm>
- [5] Rollins, N. C., et al. (2016). Why invest, and what it will take to improve breastfeeding practices? *The Lancet*, 387(10017), 491–504. Available at: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(15\)01044-2/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)01044-2/fulltext)
- [6] Muluneh, M. W. (2023). Determinants of exclusive breastfeeding practices among mothers in Ethiopia. *PLOS ONE*. Available at: <https://doi.org/10.1371/journal.pone.0281576>
- [7] Asemahagn, M. A. (2016). Determinants of exclusive breastfeeding practices among mothers in Azezo district, northwest Ethiopia. *International Breastfeeding Journal*, 11, 22. Available at: <https://doi.org/10.1186/s13006-016-0081-x>
- [8] The DHS Program (ICF). (n.d.). *DHS Methodology: Survey Design and Sampling*. Available at: <https://dhsprogram.com/methodology/Survey-Types/DHS-Methodology.cfm>
- [9] Central Statistical Agency (CSA) [Ethiopia] and ICF. (2016). *Ethiopia Demographic and Health Survey 2016: Final Report*. Available at: <https://dhsprogram.com/publications/publication-FR328-DHS-Final-Reports.cfm>
- [10] Rutstein, S. O., & Rojas, G. (2006). *Guide to DHS Statistics*. Available at: <https://dhsprogram.com/publications/publication-dhsg1-dhs-questionnaires-and-manuals.cfm>
- [11] Heeringa, S. G., West, B. T., & Berglund, P. A. (2017). *Applied Survey Data Analysis*. Available at: <https://www.routledge.com/Applied-Survey-Data-Analysis/Heeringa-West-Berglund/p/book/9781466565303>

[12] Central Statistical Agency (CSA) [Ethiopia] and ICF. (2016). Ethiopia Demographic and Health Survey 2016: Final Report. Available at:  
<https://dhsprogram.com/publications/publication-FR328-DHS-Final-Reports.cfm>

[13] Victora, C. G., et al. (2016). Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effect. *The Lancet*, 387(10017), 475–490. Available at:  
[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(15\)01024-7/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)01024-7/fulltext)

[14] Rollins, N. C., et al. (2016). Why invest, and what it will take to improve breastfeeding practices? *The Lancet*, 387(10017), 491–504. Available at:  
[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(15\)01044-2/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)01044-2/fulltext)