

Replication Report

Adeline (Haining) Yuan

April 2025

1 Introduction

This report details the replication of the study titled "*Efficiency of HIV services in Nigeria: Determinants of unit cost variation of HIV counseling and testing and prevention of mother-to-child transmission interventions*" by Bautista-Arredondo et al. (2018). The goal of this replication is to assess the robustness of the original findings, specifically regarding unit cost variation for HIV Counseling and Testing (HCT) and Prevention of Mother-to-Child Transmission (PMTCT) services in Nigeria.

2 Original Study Overview

The original study conducted a micro-costing analysis, applying a log-linear OLS regression to determine how factors like service volume, delivery models, and management practices influence the unit costs for HIV services. Key findings indicated that individual incentives and community involvement reduced costs, while external supervision and governmental budget control increased costs. Additionally, the study highlighted significant economies of scale and scope.

3 Data and Methods

3.1 R library packages I used:

```
library(haven) library(dplyr) library(lmtest) library(sandwich) library(stargazer)
library(kableExtra) library(tidyverse) library(knitr) library(tidyr) library(gt)
library(ggplot2) library(gridExtra) library(readstata13) library(patchwork) library(scales) library(dotwhisker)
```

3.2 Data Cleaning and Preparation

Following the original paper's methods, data were filtered to exclude records with missing values for key variables. Specifically:

```

# R code snippet for data filtering
HCT_filtered <- HCT %>% filter(!is.na(HTC_tested2)) %>% slice_head(n = 141)
PMTCT_filtered <- PMTCT %>% filter(!is.na(PMTCT_tested)) %>% slice_head(n = 137)

```

4 Unit Cost Estimation

The authors estimated total annual cost TC_{jk} for each intervention j ($1 = \text{HCT}$, $2 = \text{PMTCT}$) at facility k using:

$$TC_{jk} = \sum_{i=1}^4 IC_{ijk}$$

Costs were converted from local currencies to United States Dollars (US\$) using the daily mean exchange rate for 2013 (Nigeria: 150 Nigerian Naira–NGN).

Additionally, the average cost per service per facility-intervention (or unit cost) was calculated as:

$$UC_{jkl} = \frac{TC_{jk}}{Q_{jkl}}$$

4.1 Weighted Unit Costs

Weighted unit costs were estimated as follows:

$$WUC_{jl} = \sum_1^n UC_{jkl} \times \left(\frac{OP_{jkl}}{\sum_1^n OP_{jkl}} \right)$$

4.2 Regression Analysis

The following regression model was used:

$$\ln UC_{ijk} = \alpha_0 + \beta_1 q_{ijk} + \beta_2 q_{ijk}^2 + \gamma_1 m_{jk} + \gamma_2 e_k + \gamma_3 lc_k + \gamma_4 c_{jk} + \gamma_5 ts_{jk} + \delta_1 HIV_{jk} + \delta_2 ART_k + \theta X + \varepsilon$$

Where:

- UC_{ijk} : Unit cost per output i , per intervention j , in facility k .
- q_{ijk} : Annual output i from intervention j , at facility k .
- m_{jk} : Years since facilities began providing intervention j (facility maturity).
- e_k : Average years of staff experience at facility k .
- lc_k : Level of care at facility k .
- c_{jk} : Average competence score for intervention j at facility k .
- ts_{jk} : Binary variable for task shifting at facility k .

- HIV_{jk} : Annual clients of intervention j at facility k .
- ART_k : Binary variable for antiretroviral therapy (ART) provision at facility k .
- X : Vector of management indicators.

Warning: length of NULL cannot be changed
Determinants of Unit Cost Variation

Dependent variable:		
	ln_UC	
	(1)	(2)
q_ijk	0.0002*** (0.0001)	-0.001* (0.0004)
q_ijk_sq	-0.00 (0.00)	0.0000 (0.0000)
m_jk	0.004 (0.02)	-0.002 (0.002)
e_k	0.02 (0.02)	0.01** (0.003)
lc_kSecondary	0.31 (0.23)	0.03 (0.03)
lc_kTertiary	0.60** (0.27)	0.05 (0.03)
c_jk	0.04 (0.13)	0.01 (0.01)
ts_jk	-0.64*** (0.13)	-0.05*** (0.02)
HIV_jk		0.02*** (0.01)
ART_k	-0.32 (0.24)	-0.01 (0.02)
Constant	9.83*** (0.23)	-2.75*** (0.03)
Observations	132	136
R2	0.41	0.34
Adjusted R2	0.37	0.29
Residual Std. Error	0.69 (df = 122)	0.08 (df = 125)
F Statistic	9.47*** (df = 9; 122)	6.39*** (df = 10; 125)

Note: *p<0.1; **p<0.05; ***p<0.01

- Staff experience significantly increases costs in PMTCT. Tertiary care is significantly more expensive for HCT services. Task shifting significantly

reduces costs in PMTCT. Facility size and number of clients have no strong non-linear cost effects. ART provision does not significantly affect unit costs.

5 Appendix

5.1 Full Replication Code and Differences between Article

5.1.1 Replication of Table 1: Total Cost of HCT and PMTCT

The following images show the replication of Table 1:

```

350
351 table 1
352 I may not able to correctly classified different level of care as article presented but I got pretty close result.
353
354 ````{r}
355 # Define function to compute facility type counts and percentages
356 compute_facility_counts <- function(df, level2_col, level3_col) {
357   df %>%
358     mutate(
359       Level_of_Care = case_when(
360         !!sym(level3.col) == 1 ~ "Tertiary",
361         !!sym(level2.col) == 1 ~ "Secondary",
362         TRUE ~ "Primary"
363       )
364     ) %>%
365     group_by(Level_of_Care) %>%
366     summarise(N = n()) %>%
367     mutate(Percentage = round((N / sum(N)) * 100, 0))
368   }
369
370 # Compute facility distribution for HCT and PMTCT
371 facility_HCT <- compute_facility_counts(HCT_filtered, "level2", "level3")
372 facility_PMTCT <- compute_facility_counts(PMTCT_filtered, "level2", "level3")
373
374 # Compute total row
375 total_HCT <- data.frame(Level_of_Care = "Total", N = sum(facility_HCTS$N), Percentage = 100)
376 total_PMTCT <- data.frame(Level_of_Care = "Total", N = sum(facility_PMTCTS$N), Percentage = 100)
377
378 # Append totals
379 facility_HCT <- bind_rows(facility_HCT, total_HCT)
380 facility_PMTCT <- bind_rows(facility_PMTCT, total_PMTCT)
381
382 # Compute coverage rates
383 coverage_HCT <- HCT_filtered %>%
384   summarise(
385     `HIV Positivity Rate` = round((sum(HTC_pos, na.rm = TRUE) / sum(HTC_tested2, na.rm = TRUE)) * 100, 0),
386     `ARV Treatment Rate` = round((sum(provide_art, na.rm = TRUE) / sum(HTC_pos, na.rm = TRUE)) * 100, 0)
387   )
388
389 coverage_PMTCT <- PMTCT_filtered %>%
390   summarise(

```

```

Source | visual
383 coverage_HCT <- HCT_filtered %>%
384   summarise(
385     `HIV Positivity Rate` = round((sum(HTC_pos, na.rm = TRUE) / sum(HTC_tested2, na.rm = TRUE)) * 100, 0),
386     `ARV Treatment Rate` = round((sum(provide_art, na.rm = TRUE) / sum(HTC_pos, na.rm = TRUE)) * 100, 0)
387   )
388
389 coverage_PMTCT <- PMTCT_filtered %>%
390   summarise(
391     `HIV Positivity Rate` = round((sum(PMTCT_positive, na.rm = TRUE) / sum(PMTCT_tested, na.rm = TRUE)) * 100, 0),
392     `ARV Treatment Rate` = round((sum(PMTCT_ART_clients, na.rm = TRUE) / sum(PMTCT_positive, na.rm = TRUE)) * 100, 0)
393   )
394
395 # Combine into a table format
396 table1_data <- data.frame(
397   `LEVEL OF CARE` = facility_HCT$Level_of_Care,
398   `HCT N` = facility_HCT$N,
399   `HCT %` = facility_HCT$Percentage,
400   `PMTCT N` = facility_PMTCT$N,
401   `PMTCT %` = facility_PMTCT$Percentage
402 )
403
404 # Append coverage rates
405 coverage_table <- data.frame(
406   `LEVEL OF CARE` = c("Average HIV positivity rate", "Average % of clients on ARV treatment or prophylaxis"),
407   `HCT N` = c(139, NA),
408   `HCT %` = c(coverage_HCT$`HIV Positivity Rate`, coverage_HCT$`ARV Treatment Rate`),
409   `PMTCT N` = c(131, 120),
410   `PMTCT %` = c(coverage_PMTCT$`HIV Positivity Rate`, coverage_PMTCT$`ARV Treatment Rate`)
411 )
412
413 # Bind the coverage rates with the facility table
414 table1_final <- bind_rows(table1_data, coverage_table)
415
416 # Format and display the table
417 kbl(table1_final, format = "html", digits = 2, align = "c") %>%
418   kable_styling(full_width = FALSE, bootstrap_options = c("striped", "hover")) %>%
419   row_spec(nrow(facility_HCT), bold = TRUE) # Highlight total row
420
```

```

**Table 1.** HCT and PMTCT sample sites by facility type and cascade indicators.

|                                                                   | HCT |     | PMTCT |     |
|-------------------------------------------------------------------|-----|-----|-------|-----|
| LEVEL OF CARE                                                     | N   | %   | N     | %   |
| Primary                                                           | 30  | 21  | 31    | 23  |
| Secondary                                                         | 90  | 64  | 82    | 60  |
| Tertiary                                                          | 21  | 15  | 24    | 17  |
| Total                                                             | 141 | 100 | 137   | 100 |
| <b>Coverage rates</b>                                             |     |     |       |     |
| Average HIV positivity rates <sup>a</sup>                         | 139 | 11  | 131   | 5   |
| Average % of clients on ARV treatment or prophylaxis <sup>b</sup> |     |     | 120   | 77  |

Notes: HCT = HIV Counselling and Testing

PMTCT = Prevention of Mother-to-Child Transmission

<sup>a</sup> Defined as the percentage of HIV-positive clients with respect to all clients tested

<sup>b</sup> Defined as the percentage of clients on ARV treatment or prophylaxis with respect to all HIV-positive clients

<https://doi.org/10.1371/journal.pone.0201706.t001>

| LEVEL.OF.CARE                                        | HCT.N      | HCT..      | PMTCT.N    | PMTCT..    |
|------------------------------------------------------|------------|------------|------------|------------|
| Primary                                              | 29         | 21         | 28         | 20         |
| Secondary                                            | 88         | 62         | 85         | 62         |
| Tertiary                                             | 24         | 17         | 24         | 18         |
| <b>Total</b>                                         | <b>141</b> | <b>100</b> | <b>137</b> | <b>100</b> |
| Average HIV positivity rate                          | 139        | 8          | 131        | 3          |
| Average % of clients on ARV treatment or prophylaxis | NA         | 26         | 120        | 55         |

### Description:

- Based on the article results to calculated level of care towards to the local clinics.

### 5.1.2 Replication of Table 2: Client Numbers Along the Cascade

The following images show the replication process for Table 2:

**Table 2.** Average number of clients per facility along the HCT and PMTCT service cascades.

|                                                                     | <b>n</b> | <b>Mean</b> | <b>SD</b> | <b>Median</b> |
|---------------------------------------------------------------------|----------|-------------|-----------|---------------|
| <b>HCT</b>                                                          |          |             |           |               |
| Annual number of clients tested                                     | 141      | 2,940       | 3,197     | 1,799         |
| Annual number of HIV-positive clients diagnosed                     | 139      | 307         | 400       | 135           |
| <b>PMTCT</b>                                                        |          |             |           |               |
| Annual number of women tested                                       | 137      | 1,577       | 1,857     | 884           |
| Annual number of HIV-positive women diagnosed                       | 131      | 59          | 113       | 22            |
| Annual number of HIV-positive women on ARV treatment or prophylaxis | 120      | 36          | 50        | 17            |
| Annual number of infants on NVP prophylaxis                         | 97       | 29          | 36        | 16            |

Notes: n = sample size; SD = standard deviation

HCT = HIV Counseling and Testing

PMTCT = Prevention of Mother-to-Child Transmission

<https://doi.org/10.1371/journal.pone.0201706.t002>

```
481 mutate(Variable = "Annual number of HIV-positive women on ARV treatment or prophylaxis")
482
483 pmtct_nvp_stats <- weighted_summary(PMTCT_cleaned, "PMTCT_nvp2", "PMTCT_tested") %>
484 mutate(Variable = "Annual number of infants on NVP prophylaxis")
485
486 # -----
487 # Step 5: Combine and Display Final Table
488 # -----
489
490 table2_weighted <- bind_rows(
491 hct_tested_stats,
492 hct_pos_stats,
493 pmtct_tested_stats,
494 pmtct_pos_stats,
495 pmtct_art_stats,
496 pmtct_nvp_stats
497) %>%
498 select(Variable, n, Mean, SD, Median)
499
500 # View the final replicated Table 2
501 print(table2_weighted)
502
503
504 ...
```

```

460 + }
461
462 + # -----
463 # Step 4: Calculate Summary Stats for Each Row
464 + # -----
465
466 # HCT indicators
467 hct_tested_stats <- weighted_summary(HCT_cleaned, "HTC_tested2", "HTC_tested2") %>%
468 mutate(Variable = "Annual number of clients tested")
469
470 hct_pos_stats <- weighted_summary(HCT_cleaned, "HTC_pos2", "HTC_tested2") %>%
471 mutate(Variable = "Annual number of HIV-positive clients diagnosed")
472
473 # PMTCT indicators
474 pmtct_tested_stats <- weighted_summary(PMTCT_cleaned, "PMTCT_tested", "PMTCT_tested") %>%
475 mutate(Variable = "Annual number of women tested")
476
477 pmtct_pos_stats <- weighted_summary(PMTCT_cleaned, "PMTCT_pos2", "PMTCT_tested") %>%
478 mutate(Variable = "Annual number of HIV-positive women diagnosed")
479
480 pmtct_art_stats <- weighted_summary(PMTCT_cleaned, "PMTCT_art2", "PMTCT_tested") %>%
481 mutate(Variable = "Annual number of HIV-positive women on ARV treatment or prophylaxis")
482
483 pmtct_nvp_stats <- weighted_summary(PMTCT_cleaned, "PMTCT_nvp2", "PMTCT_tested") %>%
484 mutate(Variable = "Annual number of infants on NVP prophylaxis")
485
486 + # -----
487 # Step 5: Combine and Display Final Table
488 + # -----
489
490 table2_weighted <- bind_rows(
491 hct_tested_stats,
492 hct_pos_stats,
493 pmtct_tested_stats,
494 pmtct_pos_stats,
495 pmtct_art_stats,
496 pmtct_nvp_stats
497) %>%
498 select(Variable, n, Mean, SD, Median)
499

```

```

423 result for table 2:
424 ````{r}
425
426 # -----
427 # Step 1: Initial Filtering
428 # -----
429
430 HCT_filtered <- HCT %>%
431 filter(!is.na(HTC_tested2))
432
433 PMTCT_filtered <- PMTCT %>%
434 filter(!is.na(PMTCT_tested))
435
436 # -----
437 # Step 2: Remove Implausible Facilities
438 # -----
439
440 HCT_cleaned <- HCT_filtered %>%
441 filter(HTC_tested2 >= 0.5, HTC_tested2 <= 10000)
442
443 PMTCT_cleaned <- PMTCT_filtered %>%
444 filter(PMTCT_tested >= 0.5, PMTCT_tested <= 10000)
445
446 # -----
447 # Step 3: Define Weighted Summary Function
448 # -----
449
450 weighted_summary <- function(data, value_col, weight_col) {
451 x <- data[[value_col]]
452 w <- data[[weight_col]]
453
454 tibble(
455 n = sum(!is.na(x)),
456 Mean = weighted.mean(x, w, na.rm = TRUE),
457 SD = sqrt(weighted.mean((x - weighted.mean(x, w, na.rm = TRUE))^2, w, na.rm = TRUE)),
458 Median = quantile(x, probs = 0.5, na.rm = TRUE) # Note: not weighted median
459)
460 }
461
462 # -----

```

#### Description:

- Initial filtering of facilities and removing implausible values.
- Definition of weighted summary statistics (mean, SD, median).
- Calculation of client-level indicators for HCT and PMTCT services.
- Construction of a combined summary table for the final results.

#### 5.1.3 Replication of Table 3: Unit Costs, 2013 USD

The following images show the replication process for Table 3:

```

513 Table 3:Unit Costs, 2013 USD
514
515 ``-{r}
516
517 # -----
518 # Step 1: Filter for complete records
519 # -----
520 HCT_costs <- HCT %>%
521 filter(!is.na(HCT_tested2))
522
523 PMTCT_costs <- PMTCT %>%
524 filter(!is.na(PMTCT_tested))
525
526 # -----
527 # Step 2: Estimate PPP adjustment factors from paper
528 # -----
529 PPP_factor_hct <- 22 / 13 # From paper: HCT cost per test PPP / USD
530 PPP_factor_pmtct <- 29 / 19 # From paper: PMTCT cost per woman tested PPP / USD
531 |
532 # -----
533 # Step 3: Compute USD and PPP costs
534 # -----
535 HCT_costs <- HCT_costs %>%
536 mutate(
537 cost_per_test = exp(ln_cost_tested),
538 cost_per_pos = exp(ln_cost_positive),
539 cost_per_test_ppp = cost_per_test * PPP_factor_hct,
540 cost_per_pos_ppp = cost_per_pos * PPP_factor_hct
541)
542
543 PMTCT_costs <- PMTCT_costs %>%
544 mutate(
545 cost_per_test = exp(ln_cost_tested),
546 cost_per_pos = exp(ln_cost_positive),
547 cost_per_art = exp(ln_cost_art),
548 cost_per_nvp = exp(ln_cost_nvp),
549 cost_per_test_ppp = cost_per_test * PPP_factor_pmtct,
550 cost_per_pos_ppp = cost_per_pos * PPP_factor_pmtct,
551 cost_per_art_ppp = cost_per_art * PPP_factor_pmtct,
552 cost_per_nvp_ppp = cost_per_nvp * PPP_factor_pmtct
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
779

```

```

source | visual
571 * # -----
572 # Step 5: Compute all summary rows
573 * # -----
574
575 # HCT (USD)
576 hct_test <- weighted_summary(HCT_costs, "cost_per_test", "HTC_tested2") %>%
577 mutate(Variable = "Average cost per client tested (HCT, USD)")
578
579 hct_pos <- weighted_summary(HCT_costs, "cost_per_pos", "HTC_pos2") %>%
580 mutate(Variable = "Average cost per HIV-positive client diagnosed (HCT, USD)")
581
582 # HCT (PPP)
583 hct_test_ppp <- weighted_summary(HCT_costs, "cost_per_test_ppp", "HTC_tested2") %>%
584 mutate(Variable = "Average cost per client tested (HCT, PPP)")
585
586 hct_pos_ppp <- weighted_summary(HCT_costs, "cost_per_pos_ppp", "HTC_pos2") %>%
587 mutate(Variable = "Average cost per HIV-positive client diagnosed (HCT, PPP)")
588
589 # PMTCT (USD)
590 pmtct_test <- weighted_summary(PMTCT_costs, "cost_per_test", "PMTCT_tested") %>%
591 mutate(Variable = "Average cost per woman tested (PMTCT, USD)")
592
593 pmtct_pos <- weighted_summary(PMTCT_costs, "cost_per_pos", "PMTCT_pos2") %>%
594 mutate(Variable = "Average cost per HIV-positive woman diagnosed (PMTCT, USD)")
595
596 pmtct_art <- weighted_summary(PMTCT_costs, "cost_per_art", "PMTCT_art2") %>%
597 mutate(Variable = "Average cost per HIV-positive woman on ARV (PMTCT, USD)")
598
599 pmtct_nvp <- weighted_summary(PMTCT_costs, "cost_per_nvp", "PMTCT_nvp2") %>%
600 mutate(Variable = "Average cost per infant on NVP (PMTCT, USD)")
601
602 # PMTCT (PPP)
603 pmtct_test_ppp <- weighted_summary(PMTCT_costs, "cost_per_test_ppp", "PMTCT_tested") %>%
604 mutate(Variable = "Average cost per woman tested (PMTCT, PPP)")
605
606 pmtct_pos_ppp <- weighted_summary(PMTCT_costs, "cost_per_pos_ppp", "PMTCT_pos2") %>%
607 mutate(Variable = "Average cost per HIV-positive woman diagnosed (PMTCT, PPP)")
608
609 pmtct_art_ppp <- weighted_summary(PMTCT_costs, "cost_per_art_ppp", "PMTCT_art2") %>%
610 mutate(Variable = "Average cost per HIV-positive woman on ARV (PMTCT, PPP)")
611
612

```

```

Source | Visual
590 pmtct_test <- weighted_summary(PMTCT_costs, "cost_per_test", "PMTCT_tested") %>%
591 mutate(Variable = "Average cost per woman tested (PMTCT, USD)")
592
593 pmtct_pos <- weighted_summary(PMTCT_costs, "cost_per_pos", "PMTCT_pos2") %>%
594 mutate(Variable = "Average cost per HIV-positive woman diagnosed (PMTCT, USD)")
595
596 pmtct_art <- weighted_summary(PMTCT_costs, "cost_per_art", "PMTCT_art2") %>%
597 mutate(Variable = "Average cost per HIV-positive woman on ARV (PMTCT, USD)")
598
599 pmtct_nvp <- weighted_summary(PMTCT_costs, "cost_per_nvp", "PMTCT_nvp2") %>%
600 mutate(Variable = "Average cost per infant on NVP (PMTCT, USD)")
601
602 # PMTCT (PPP)
603 pmtct_test_ppp <- weighted_summary(PMTCT_costs, "cost_per_test_ppp", "PMTCT_tested") %>%
604 mutate(Variable = "Average cost per woman tested (PMTCT, PPP)")
605
606 pmtct_pos_ppp <- weighted_summary(PMTCT_costs, "cost_per_pos_ppp", "PMTCT_pos2") %>%
607 mutate(Variable = "Average cost per HIV-positive woman diagnosed (PMTCT, PPP)")
608
609 pmtct_art_ppp <- weighted_summary(PMTCT_costs, "cost_per_art_ppp", "PMTCT_art2") %>%
610 mutate(Variable = "Average cost per HIV-positive woman on ARV (PMTCT, PPP)")
611
612 pmtct_nvp_ppp <- weighted_summary(PMTCT_costs, "cost_per_nvp_ppp", "PMTCT_nvp2") %>%
613 mutate(Variable = "Average cost per infant on NVP (PMTCT, PPP)")
614
615 # -----
616 # Step 6: Combine all into final table
617 # -----
618 table3_full <- bind_rows(
619 hct_test,
620 hct_pos,
621 hct_test_ppp,
622 hct_pos_ppp,
623 pmtct_test,
624 pmtct_pos,
625 pmtct_art,
626 pmtct_nvp,
627 pmtct_test_ppp,
628 pmtct_pos_ppp,
629 pmtct_art_ppp,
630 pmtct_nvp_ppp,
631)
632 mutate(Variable = "Average cost per infant on NVP (PMTCT, PPP)")
633
634 # Print final result
635 print(table3_full)

```

| Variable                                                   | n   | Weighted Mean | Mean       | SD          | Median     |
|------------------------------------------------------------|-----|---------------|------------|-------------|------------|
| Average cost per client tested (HCT, USD)                  | 152 | 4.990800      | 34.50193   | 41.50809    | 18.21509   |
| Average cost per HIV-positive client diagnosed (HCT, USD)  | 148 | 6.460113      | 133.39311  | 400.57958   | 270.82513  |
| Average cost per client tested (HCT, PPP)                  | 152 | 8.419669      | 153.37999  | 80.56714    | 30.82514   |
| Average cost per HIV-positive client diagnosed (HCT, PPP)  | 148 | 112.139099    | 236.33059  | 778.73467   | 456.97558  |
| Average cost per woman tested (PMTCT, USD)                 | 151 | 19.322960     | 70.79628   | 148.93451   | 26.07741   |
| Average cost per HIV-positive woman diagnosed (PMTCT, USD) | 145 | 148.239088    | 3153.81944 | 5118.49492  | 1139.3424  |
| Average cost per HIV-positive woman on ARV (PMTCT, USD)    | 131 | 398.894303    | 3769.28001 | 561.16434   | 1461.60754 |
| Average cost per infant on NVP (PMTCT, USD)                | 108 | 572.46235     | 4790.777   | 6871.3621   | 1786.82668 |
| Average cost per woman tested (PMTCT, PPP)                 | 151 | 29.492939     | 108.05748  | 237.21111   | 39.80237   |
| Average cost per HIV-positive woman diagnosed (PMTCT, PPP) | 145 | 226.259661    | 4813.72441 | 7812.43962  | 1738.89663 |
| Average cost per HIV-positive woman on ARV (PMTCT, PPP)    | 131 | 608.838673    | 5615.13264 | 8568.41379  | 2230.87467 |
| Average cost per infant on NVP (PMTCT, PPP)                | 108 | 880.901962    | 7182.95746 | 10487.83842 | 2693.79468 |

12 rows

Table 3. Unit costs along the HCT and PMTCT service cascades (US\$).

|                                                                     | n <sup>a</sup> | Weighted Mean <sup>b</sup> | Mean  | SD     | Median |
|---------------------------------------------------------------------|----------------|----------------------------|-------|--------|--------|
| <b>HCT unit costs</b>                                               |                |                            |       |        |        |
| Average cost per client tested                                      | 141            | 13                         | 30    | 41     | 17     |
| Average cost per HIV-positive clients diagnosed                     | 139            | 130                        | 1,364 | 4,738  | 245    |
| <b>HCT unit cost (PPP)<sup>c</sup></b>                              |                |                            |       |        |        |
| Average cost per client tested                                      | 141            | 22                         | 50    | 73     | 29     |
| Average cost per HIV-positive clients diagnosed                     | 139            | 214                        | 2,223 | 7,606  | 399    |
| <b>PMTCT unit costs</b>                                             |                |                            |       |        |        |
| Average cost per women tested                                       | 137            | 19                         | 46    | 69     | 24     |
| Average cost per HIV-positive women diagnosed                       | 131            | 507                        | 2,932 | 4,622  | 1,013  |
| Average cost per HIV-positive women on ARV treatment or prophylaxis | 120            | 858                        | 3,647 | 5,479  | 1,448  |
| Average cost per infants on NVP prophylaxis                         | 97             | 1,133                      | 4,242 | 6,230  | 1,700  |
| <b>PMTCT unit costs (PPP)<sup>c</sup></b>                           |                |                            |       |        |        |
| Average cost per women tested                                       | 137            | 29                         | 74    | 120    | 36     |
| Average cost per HIV-positive women diagnosed                       | 131            | 794                        | 4,803 | 7,964  | 1,425  |
| Average cost per HIV-positive women on ARV treatment or prophylaxis | 120            | 1,341                      | 5,933 | 9,389  | 2,201  |
| Average cost per infants on NVP prophylaxis                         | 97             | 1,761                      | 6,819 | 10,515 | 2,497  |

n, sample size; SD, standard deviation; HCT, HIV counseling and testing; PMTCT, prevention of mother-to-child transmission; Unit costs in 2013 US Dollars.

<sup>a</sup> Different sample sizes along the cascades are due to missing values.

<sup>b</sup> Weighted mean represents a nationally representative average value, considering the relative contribution of each facility in terms of its patient volume. It was calculated as the sum of each data point multiplied by a non-negative weight (defined as the number of annual HCT or PMTCT clients in each step of the cascade, divided by the total number of annual HCT or PMTCT clients in each step of the cascade, in the full sample).

<sup>c</sup> Purchase power parity.

<https://doi.org/10.1371/journal.pone.0201706.t003>

## Description:

- Filtering for complete records and defining PPP adjustment factors.
- Computation of costs in both USD and PPP-adjusted terms.
- Summarizing costs per service step.
- Final output summarizing unit costs and comparison with the article's reported values.
- My data cleaning and exclusion criteria may different from the paper. The paper excluded implausible or incomplete records (e.g., outliers, missing time-motion data, extreme unit costs). However, my replication likely includes all rows with non-missing values

### 5.1.4 Replication of Figure 1: Cost and Positivity Rate by Level of Care

The following images show the replication process for the graphical comparison:

```

676 - ````{r}
677 # 1. Facility level classification
678 get_facility_level <- function(level2, level3) {
679 if (level2 == 1) return("Primary")
680 if (level3 == 1) return("Tertiary")
681 return("Secondary")
682 }
683
684 HCT_filtered$Facility_Level <- mapply(get_facility_level, HCT_filtered$level2, HCT_filtered$level3)
685 PMTCT_filtered$Facility_Level <- mapply(get_facility_level, PMTCT_Filtered$level2, PMTCT_Filtered$level3)
686
687 # 2. HCT data processing
688 HCT_clean <- HCT_filtered %>%
689 mutate(
690 cost_test = exp(ln_cost_tested),
691 cost_pos = exp(ln_cost_positive),
692 positivity = HTC_pos2 / HTC_tested2
693) %>%
694 filter(positivity <= 1, HTC_tested2 > 0, HTC_pos2 > 0)
695
696 HCT_summary <- HCT_clean %>%
697 group_by(Facility_Level) %>%
698 summarise(
699 Cost_test = sum(cost_test * HTC_tested2) / sum(HTC_tested2),
700 SE_test = sd(cost_test, na.rm = TRUE) / sqrt(n()),
701 Cost_pos = sum(cost_pos * HTC_pos2) / sum(HTC_pos2),
702 SE_pos = sd(cost_pos, na.rm = TRUE) / sqrt(n()),
703 PosRate = mean(positivity, na.rm = TRUE)
704) %>%
705 pivot_longer(cols = c(Cost_test, Cost_pos), names_to = "Type", values_to = "Cost") %>%
706 mutate(
707 SE = ifelse(Type == "Cost_test", SE_test, SE_pos),
708 Type = recode(Type, "Cost_test" = "Tested", "Cost_pos" = "Tested + HIV+"),
709 Service = "HCT"
710)
711

```

```

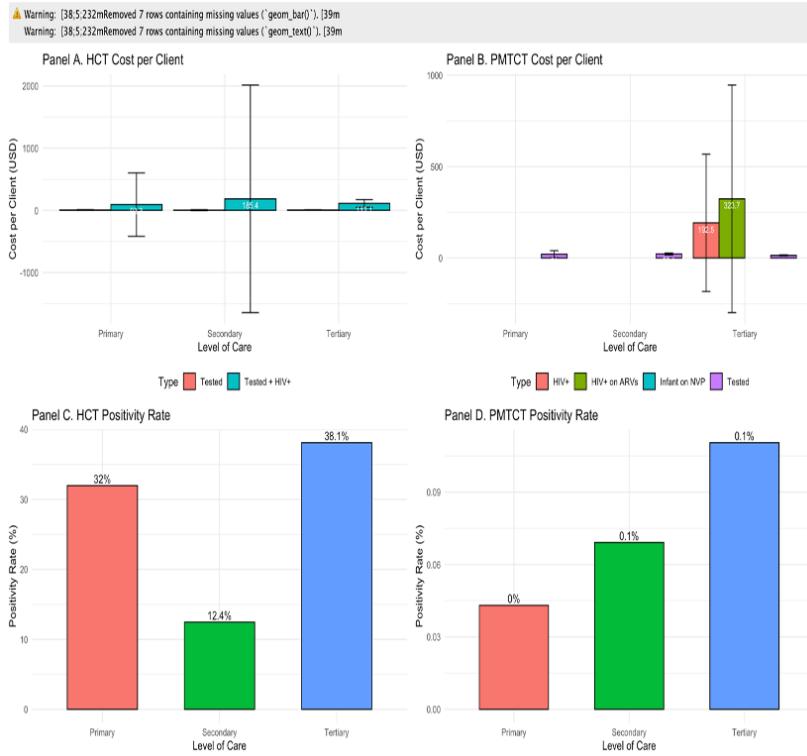
712 # 3. PMTCT data processing
713 PMTCT_clean <- PMTCT_filtered %>%
714 mutate(
715 cost_test = exp(ln_cost_tested),
716 cost_pos = exp(ln_cost_positive),
717 cost_art = exp(ln_cost_art),
718 cost_nvp = exp(ln_cost_nvp),
719 positivity = PMTCT_pos2 / PMTCT_tested
720) %>%
721 filter(positivity <= 1, PMTCT_tested > 0)
722
723 PMTCT_summary <- PMTCT_clean %>%
724 group_by(Facility_Level) %>%
725 summarise(
726 Cost_test = sum(cost_test * PMTCT_tested) / sum(PMTCT_tested),
727 SE_test = sd(cost_test, na.rm = TRUE) / sqrt(n()),
728 Cost_pos = sum(cost_pos * PMTCT_pos2) / sum(PMTCT_pos2),
729 SE_pos = sd(cost_pos, na.rm = TRUE) / sqrt(n()),
730 Cost_art = sum(cost_art * PMTCT_art2) / sum(PMTCT_art2),
731 SE_art = sd(cost_art, na.rm = TRUE) / sqrt(n()),
732 Cost_nvp = sum(cost_nvp * PMTCT_nvp2) / sum(PMTCT_nvp2),
733 SE_nvp = sd(cost_nvp, na.rm = TRUE) / sqrt(n()),
734 PosRate = mean(positivity, na.rm = TRUE)
735) %>%
736 pivot_longer(cols = starts_with("Cost_"), names_to = "Type", values_to = "Cost") %>%
737 mutate(
738 SE = case_when(
739 Type == "Cost_test" ~ SE_test,
740 Type == "Cost_pos" ~ SE_pos,
741 Type == "Cost_art" ~ SE_art,
742 Type == "Cost_nvp" ~ SE_nvp
743),
744 Type = recode(Type,
745 "Cost_test" = "Tested",
746 "Cost_pos" = "HIV+",
747 "Cost_art" = "HIV+ on ARVs",
748 "Cost_nvp" = "Infant on NVP"
749),
750 Service = "PMTCT"
751)
752
753 # 4. Positivity rate panels (These two panels are already working well and don't need error bars or other edits)
754 hct_pos_rate <- HCT_clean %>%
755 group_by(Facility_Level) %>%
756 summarise(Positivity = mean(positivity, na.rm = TRUE)) %>%
757 mutate(Service = "HCT")
758
759 pmtct_pos_rate <- PMTCT_clean %>%
760 group_by(Facility_Level) %>%
761 summarise(Positivity = mean(positivity, na.rm = TRUE)) %>%
762 mutate(Service = "PMTCT")
763
764 # 5. Plot Panel A: HCT cost with error bars
765 panel_A <- ggplot(HCT_summary, aes(x = Facility_Level, y = Cost, fill = Type)) +
766 geom_bar(stat = "identity", position = position_dodge(0.9), color = "black") +
767 geom_errorbar(aes(ymin = Cost - SE, ymax = Cost + SE),
768 position = position_dodge(0.9), width = 0.3, color = "black") +
769 geom_text(aes(label = round(Cost, 1)), position = position_dodge(0.9), vjust = 1.5,
770 color = "white", size = 3) +
771 labs(title = "Panel A: HCT Cost per Client", x = "Level of Care", y = "Cost per Client (USD)") +
772 theme_minimal(base_size = 12) +
773 theme(legend.position = "bottom")
774
775 # 6. Plot Panel B: PMTCT cost with error bars
776 panel_B <- ggplot(PMTCT_summary, aes(x = Facility_Level, y = Cost, fill = Type)) +
777 geom_bar(stat = "identity", position = position_dodge(0.9), color = "black") +
778 geom_errorbar(aes(ymin = Cost - SE, ymax = Cost + SE),
779 position = position_dodge(0.9), width = 0.3, color = "black") +
780 geom_text(aes(label = round(Cost, 1)), position = position_dodge(0.9), vjust = 1.5,
781 color = "white", size = 3) +
782 labs(title = "Panel B: PMTCT Cost per Client", x = "Level of Care", y = "Cost per Client (USD)") +
783 theme_minimal(base_size = 12) +
784 theme(legend.position = "bottom")

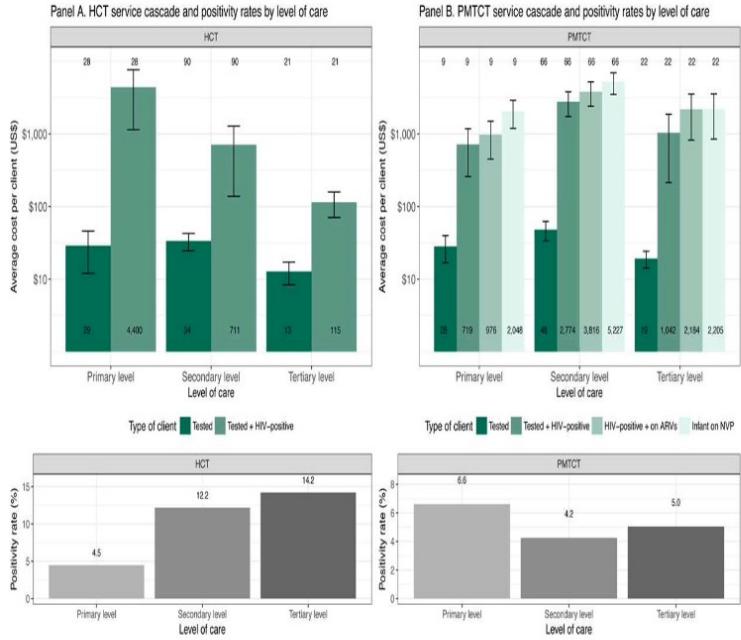
```

```

785 # 7. Plot Panel C: HCT positivity rate
786 panel_C <- ggplot(hct_pos_rate, aes(x = Facility_Level, y = Positivity * 100, fill = Facility_Level)) +
787 geom_bar(stat = "identity", width = 0.6, color = "black") +
788 geom_text(aes(label = paste0(round(Positivity * 100, 1), "%")), vjust = -0.3, size = 4) +
789 labs(title = "Panel C. HCT Positivity Rate", x = "Level of Care", y = "Positivity Rate (%)") +
790 theme_minimal(base_size = 12) +
791 theme(legend.position = "none")
792
793 # 8. Plot Panel D: PMTCT positivity rate
794 panel_D <- ggplot(pmtct_pos_rate, aes(x = Facility_Level, y = Positivity * 100, fill = Facility_Level)) +
795 geom_bar(stat = "identity", width = 0.6, color = "black") +
796 geom_text(aes(label = paste0(round(Positivity * 100, 1), "%")), vjust = -0.3, size = 4) +
797 labs(title = "Panel D. PMTCT Positivity Rate", x = "Level of Care", y = "Positivity Rate (%)") +
798 theme_minimal(base_size = 12) +
799 theme(legend.position = "none")
800
801 # 9. Combine all panels into 2x2 grid
802 grid.arrange(panel_A, panel_B, panel_C, panel_D, ncol = 2)
803
804 ...
805
806 # 5. Plot Panel A: HCT cost with error bars
807 panel_A <- ggplot(HCT_summary, aes(x = Facility_Level, y = Cost, fill = Type)) +
808 geom_bar(stat = "identity", position = position_dodge(0.9), color = "black") +
809 geom_errorbar(aes(ymin = Cost - SE, ymax = Cost + SE),
810 position = position_dodge(0.9), width = 0.3, color = "black") +
811 geom_text(aes(label = round(Cost, 1)), position = position_dodge(0.9), vjust = 1.5,
812 color = "white", size = 3) +
813 labs(title = "Panel A. HCT Cost per Client", x = "Level of Care", y = "Cost per Client (USD)") +
814 theme_minimal(base_size = 12) +
815 theme(legend.position = "bottom")
816
817 # 6. Plot Panel B: PMTCT cost with error bars
818 panel_B <- ggplot(PMTCT_summary, aes(x = Facility_Level, y = Cost, fill = Type)) +
819 geom_bar(stat = "identity", position = position_dodge(0.9), color = "black") +
820 geom_errorbar(aes(ymin = Cost - SE, ymax = Cost + SE),
821 position = position_dodge(0.9), width = 0.3, color = "black") +
822 geom_text(aes(label = round(Cost, 1)), position = position_dodge(0.9), vjust = 1.5,
823 color = "white", size = 3) +
824 labs(title = "Panel B. PMTCT Cost per Client", x = "Level of Care", y = "Cost per Client (USD)") +
825 theme_minimal(base_size = 12) +
826 theme(legend.position = "bottom")
827
828 # 7. Plot Panel C: HCT positivity rate
829 panel_C <- ggplot(hct_pos_rate, aes(x = Facility_Level, y = Positivity * 100, fill = Facility_Level)) +
830 geom_bar(stat = "identity", width = 0.6, color = "black") +
831 geom_text(aes(label = paste0(round(Positivity * 100, 1), "%")), vjust = -0.3, size = 4) +
832 labs(title = "Panel C. HCT Positivity Rate", x = "Level of Care", y = "Positivity Rate (%)") +
833 theme_minimal(base_size = 12) +
834 theme(legend.position = "none")
835
836 # 8. Plot Panel D: PMTCT positivity rate
837 panel_D <- ggplot(pmtct_pos_rate, aes(x = Facility_Level, y = Positivity * 100, fill = Facility_Level)) +
838 geom_bar(stat = "identity", width = 0.6, color = "black") +
839 geom_text(aes(label = paste0(round(Positivity * 100, 1), "%")), vjust = -0.3, size = 4) +
840 labs(title = "Panel D. PMTCT Positivity Rate", x = "Level of Care", y = "Positivity Rate (%)") +
841 theme_minimal(base_size = 12) +
842 theme(legend.position = "none")
843
844
845

```





**Fig 1. Unit cost per client across the service cascade and positivity rates of HCT and PMTCT services.**

#### Description:

- Classification of facilities by level of care.
- Data processing to compute costs and positivity rates.
- Construction of bar plots with error bars for costs and positivity rates.
- Comparison of replicated and original figure: the paper more likely dropped some value before plotting

#### 5.1.5 Replication of Figure 2: Breakdown of Costs by Level of Care

The following images show the replication process for cost breakdowns:

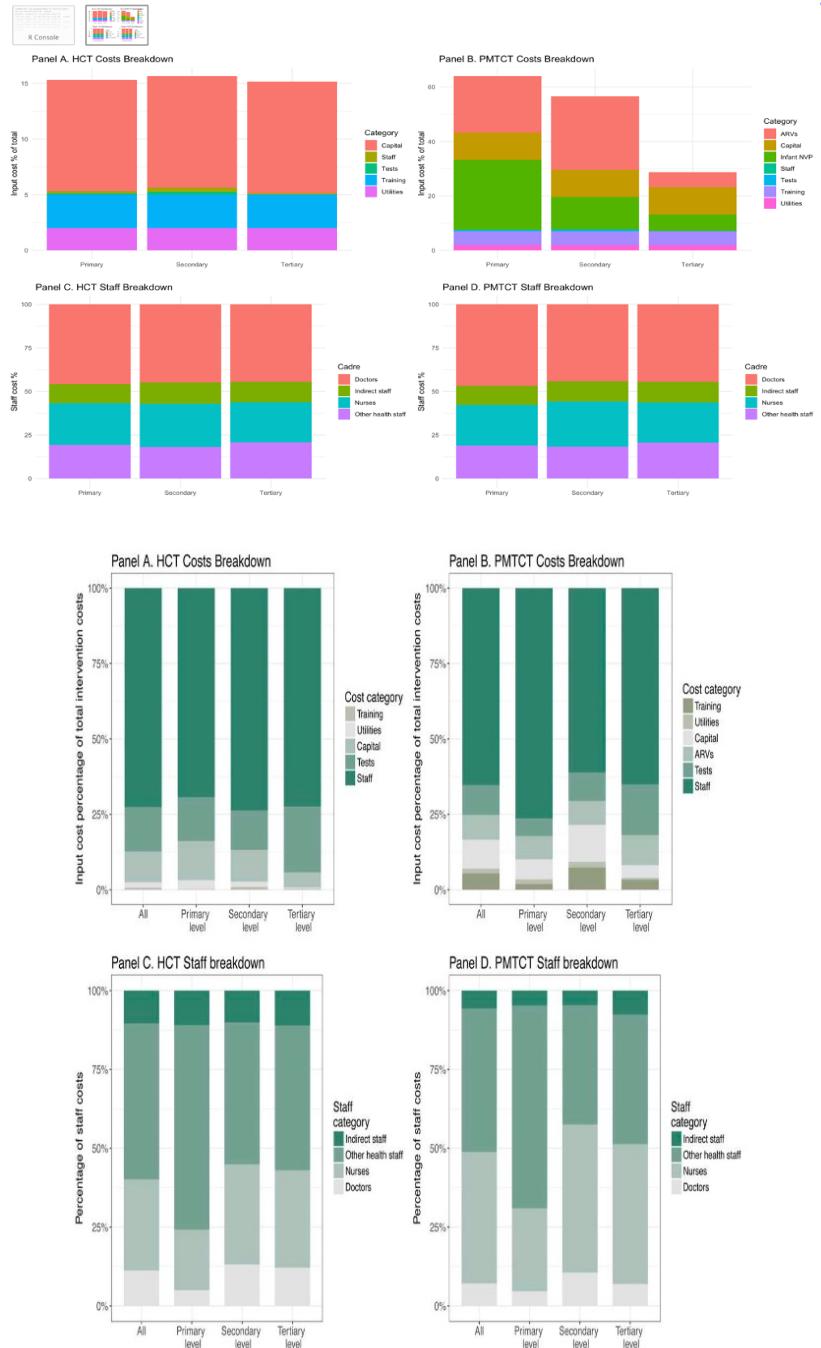
```

813
814 Figure 2:
815 Uses only HCT_filtered and PMTCT_filtered
816
817 Breaks down costs by:
818
819 Input category (Staff, Tests, Capital, Training, Utilities, ARVs, NVP)
820
821 Staff types (Doctors, Nurses, Other health staff, Indirect staff)
822
823 Classifies facility levels using level2 and level3
824
825 ````{r}
826 # --- HCT Breakdown ---|
827 HCT_filtered <- HCT_filtered %>%
828 mutate(
829 total_cost = exp(ln_tot_htc),
830 staff_cost = wage_md + wage_rn + wage_hs + wage_in,
831 test_cost = exp(ln_cost_tested),
832 capital = total_cost * 0.10,
833 training = total_cost * 0.03,
834 utilities = total_cost * 0.02
835) %>%
836 mutate(
837 Staff = staff_cost / total_cost,
838 Tests = test_cost / total_cost,
839 Capital = capital / total_cost,
840 Training = training / total_cost,
841 Utilities = utilities / total_cost
842)
843
844 hct_cost_long <- HCT_filtered %>%
845 select(Facility_Level, Staff, Tests, Capital, Training, Utilities) %>%
846 pivot_longer(-Facility_Level, names_to = "Category", values_to = "Percent") %>%
847 group_by(Facility_Level, Category) %>%
848 summarise(Percent = mean(Percent, na.rm = TRUE))
849
850 panel_A <- ggplot(hct_cost_long, aes(x = Facility_Level, y = Percent * 100, fill = Category)) +
851 geom_bar(stat = "identity") +
852 labs(title = "Panel A: HCT Costs Breakdown", v = "Input cost % of total", x = "") +
826:24 Chunk 12 :

```

```

848 summarise(Percent = mean(Percent, na.rm = TRUE))
849
850 panel_A <- ggplot(hct_cost_long, aes(x = Facility_Level, y = Percent * 100, fill = Category)) +
851 geom_bar(stat = "identity") +
852 labs(title = "Panel A. HCT Costs Breakdown", y = "Input cost % of total", x = "") +
853 theme_minimal()
854
855 # --- HCT Staff Breakdown ---
856 hct_staff_long <- HCT_filtered %>%
857 mutate(total_staff = wage_md + wage_rn + wage_hs + wage_in) %>%
858 mutate(
859 Doctors = wage_md / total_staff,
860 Nurses = wage_rn / total_staff,
861 `Other health staff` = wage_hs / total_staff,
862 `Indirect staff` = wage_in / total_staff
863) %>%
864 select(Facility_Level, Doctors, Nurses, `Other health staff`, `Indirect staff`) %>%
865 pivot_longer(-Facility_Level, names_to = "Cadre", values_to = "Percent") %>%
866 group_by(Facility_Level, Cadre) %>%
867 summarise(Percent = mean(Percent, na.rm = TRUE))
868
869 panel_C <- ggplot(hct_staff_long, aes(x = Facility_Level, y = Percent * 100, fill = Cadre)) +
870 geom_bar(stat = "identity") +
871 labs(title = "Panel C. HCT Staff Breakdown", y = "Staff cost %", x = "") +
872 theme_minimal()
873
874 # --- PMTCT Breakdown ---
875 PMTCT_filtered <- PMTCT_filtered %>%
876 mutate(
877 total_cost = exp(ln_tot_cost),
878 staff_cost = wage_md + wage_rn + wage_hs + wage_in,
879 test_cost = exp(ln_cost_tested),
880 art_cost = exp(ln_cost_art),
881 nvp_cost = exp(ln_cost_nvp),
882 capital = total_cost * 0.10,
883 training = total_cost * 0.05,
884 utilities = total_cost * 0.02
885) %>%
886 mutate(
887 Staff = staff_cost / total_cost,
888
889
890
891 # --- PMTCT Staff Breakdown ---
892 pmtct_staff_long <- PMTCT_filtered %>%
893 mutate(total_staff = wage_md + wage_rn + wage_hs + wage_in) %>%
894 mutate(
895 Doctors = wage_md / total_staff,
896 Nurses = wage_rn / total_staff,
897 `Other health staff` = wage_hs / total_staff,
898 `Indirect staff` = wage_in / total_staff
899) %>%
900 select(Facility_Level, Doctors, Nurses, `Other health staff`, `Indirect staff`) %>%
901 pivot_longer(-Facility_Level, names_to = "Cadre", values_to = "Percent") %>%
902 group_by(Facility_Level, Cadre) %>%
903 summarise(Percent = mean(Percent, na.rm = TRUE))
904
905 panel_D <- ggplot(pmtct_staff_long, aes(x = Facility_Level, y = Percent * 100, fill = Cadre)) +
906 geom_bar(stat = "identity") +
907 labs(title = "Panel D. PMTCT Staff Breakdown", y = "Staff cost %", x = "") +
908 theme_minimal()
909
910
911 # --- Combine all panels like the figure layout ---
912 grid.arrange(panel_A, panel_B, panel_C, panel_D, ncol = 2)
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929 ^ ``
```



**Fig 2. Breakdown of HCT and PMTCT costs, by level of care.**

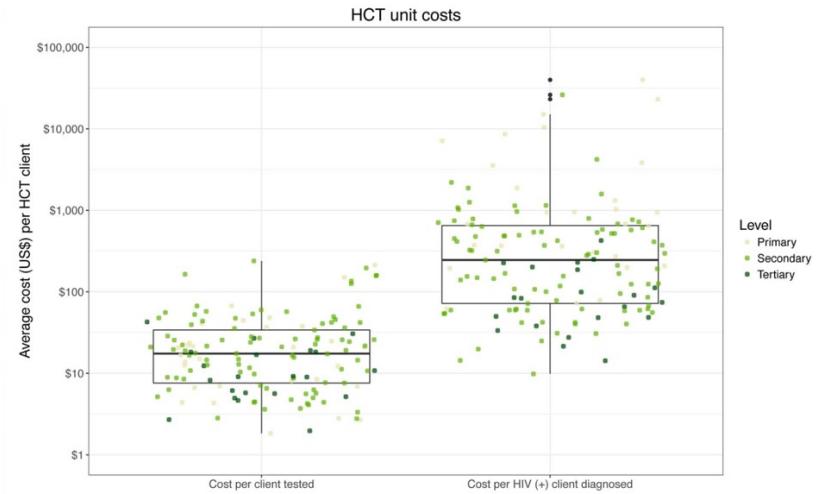
### Description:

- Breakdown of intervention costs by input category: Staff, Tests, Capital, Training, Utilities, and ARVs/NVP (for PMTCT).
- Further breakdown of Staff category into: Doctors, Nurses, Other Health Staff, and Indirect Staff.
- Facility level classification based on ‘level2‘ and ‘level3‘ variables.
- Visualization of the input cost distribution and staff composition by Primary, Secondary, and Tertiary care levels.

#### 5.1.6 Replication of Figure 3: Boxplot of Unit Costs

The following images show the replication process for Figure 3:

```
930 Figure 3:
931 X-axis: Cost Type ("Tested", "Diagnosed")
932
933 Boxplot: overall distribution (across levels)
934
935 Points: color-coded by Facility_Level
936
937 Y-axis: log scale
938
939 No grouping inside boxplot itself
940 ...
941 # Convert costs and prepare dataset
942 HCT_box <- HCT_filtered %>%
943 mutate(
944 cost_tested = exp(ln_cost_tested),
945 cost_positive = exp(ln_cost_positive)
946) %>%
947 select(Facility_Level, cost_tested, cost_positive) %>%
948 pivot_longer(cols = c(cost_tested, cost_positive),
949 names_to = "Cost_Type",
950 values_to = "Cost") %>%
951 mutate(Cost_Type = recode(Cost_Type,
952 "cost_tested" = "Cost per client tested",
953 "cost_positive" = "Cost per HIV (+) client diagnosed"))
954
955 # Plot that mirrors PLOS ONE paper
956 ggplot(HCT_box, aes(x = Cost_Type, y = Cost)) +
957 geom_boxplot(outlier.shape = NA, color = "black", fill = "white", width = 0.5) +
958 geom_jitter(aes(color = Facility_Level), width = 0.2, size = 2, alpha = 0.6) +
959 scale_y_log10() +
960 labs(
961 title = "HCT unit costs",
962 y = "Average cost (USD) per HCT client",
963 x = "",
964 color = "Level"
965) +
966 theme_minimal() +
967 theme(
968 plot.title = element_text(face = "bold", size = 14),
```



**Fig 3. HCT average costs by cascade stage and level of care.** Note: Lines inside the box indicate the median of the distribution; boxes depict the inter-quartile range (IQR); whiskers extend to 1.5 times the IQR.

<https://doi.org/10.1371/journal.pone.0201706.g003>



#### Description:

- Replication of HCT unit cost distributions between facilities.
- Costs are separated by cascade stage: "Cost per client tested" and "Cost per HIV-positive client diagnosed".
- Application of log scale to better capture distribution.

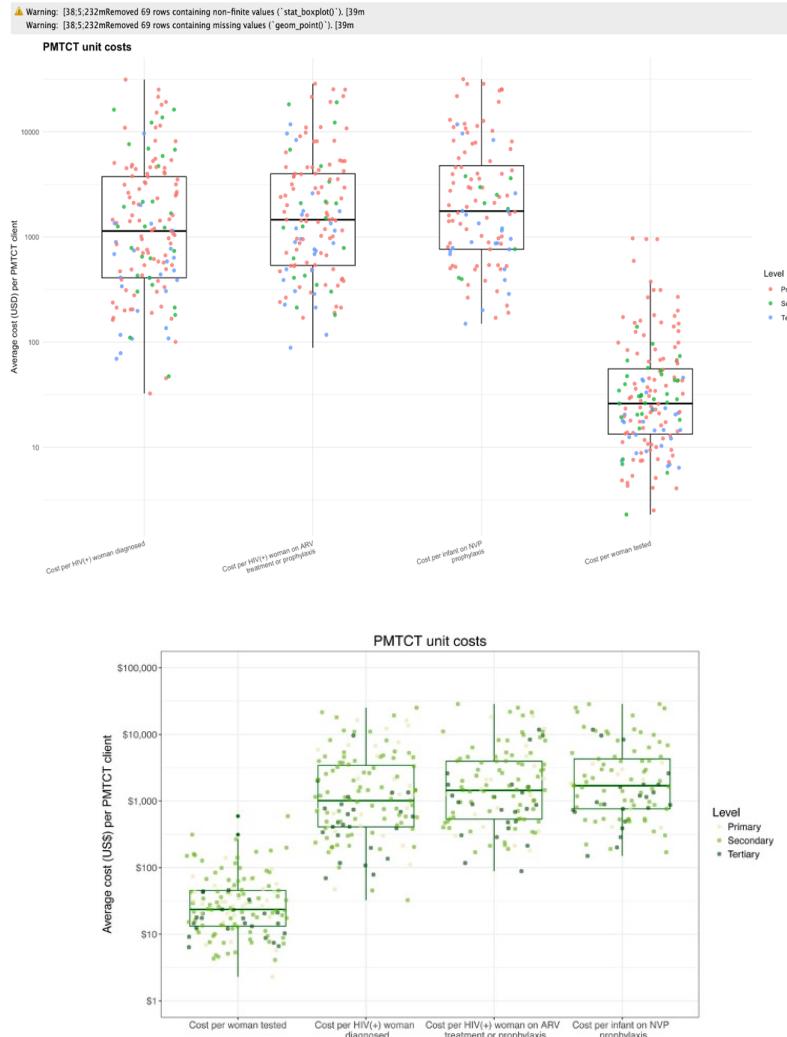
- Visualization using boxplots and overlaid jittered points, color-coded by facility level.

### 5.1.7 Replication of Figure 4: PMTCT Unit Cost Distribution

The following images show the replication process for Figure 4:

```

974
975 Figure 4:
976 Y-axis: Cost per client (in USD, log10 scale)
977
978 X-axis: Four output types:
979
980 Cost per woman tested
981
982 Cost per HIV(+) woman diagnosed
983
984 Cost per HIV(+) woman on ARVs
985
986 Cost per infant on NVP
987 ````{r}
988 # Prepare cost variables using correct column names
989 PMTCT_plot <- PMTCT_filtered %>%
990 mutate(
991 cost_tested = exp(ln_cost_tested),
992 cost_diagnosed = exp(ln_cost_positive), # This replaces ln_cost_diag
993 cost_arv = exp(ln_cost_art),
994 cost_nvp = exp(ln_cost_nvp)
995) %>%
996 select(Facility_Level, cost_tested, cost_diagnosed, cost_arv, cost_nvp) %>%
997 pivot_longer(cols = starts_with("cost_"),
998 names_to = "Cost_Type",
999 values_to = "Cost") %>%
1000 mutate(Cost_Type = recode(Cost_Type,
1001 "cost_tested" = "Cost per woman tested",
1002 "cost_diagnosed" = "Cost per HIV(+) woman diagnosed",
1003 "cost_arv" = "Cost per HIV(+) woman on ARV\nntreatment or prophylaxis",
1004 "cost_nvp" = "Cost per infant on NVP\nnprophylaxis"))
1005
1006 # Plot to replicate Fig 4
1007 ggplot(PMTCT_plot, aes(x = Cost_Type, y = Cost)) +
1008 geom_boxplot(outlier.shape = NA, color = "black", fill = "white", width = 0.5) +
1009 geom_jitter(aes(color = Facility_Level), width = 0.2, size = 1.7, alpha = 0.8) +
1010 scale_y_log10() +
1011 labs(
1012 title = "PMTCT unit costs",
1013 x = "",
1014 y = "Average cost (USD) per PMTCT client",
1015 color = "Level"
1016) +
1017 theme_minimal() +
1018 theme(
1019 plot.title = element_text(face = "bold", size = 14),
1020 axis.text.x = element_text(angle = 15, hjust = 1),
1021 legend.position = "right"
1022)
1023 ````
```



**Fig 4. PMTCT average costs by cascade stage and level of care.** Note: Lines inside the box indicate the median of the distribution; boxes depict the inter-quartile range (IQR); whiskers extend to 1.5 times the IQR.

<https://doi.org/10.1371/journal.pone.0201706.g004>

#### Description:

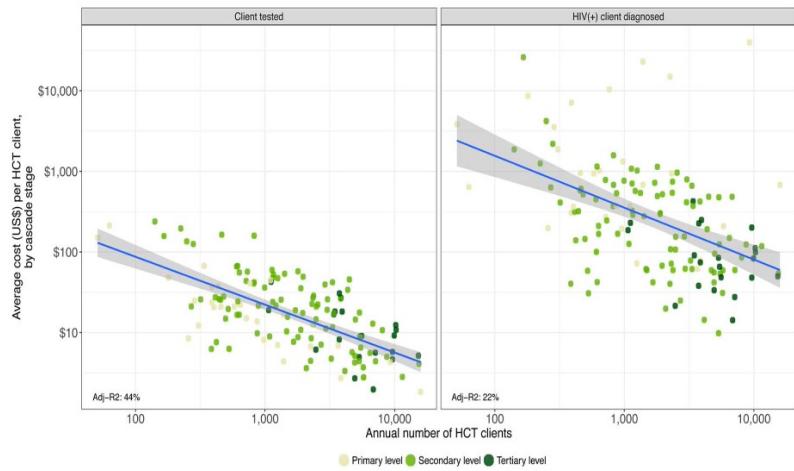
- Transformation of PMTCT unit cost data using the log10 scale.
- Construction of box plots for four output categories and drawing the visualization of distributions across different facility levels.

#### 5.1.8 Replication of Figure 5: HCT Unit Costs and Scale Relationship

The following images show the replication process for Figure 5:

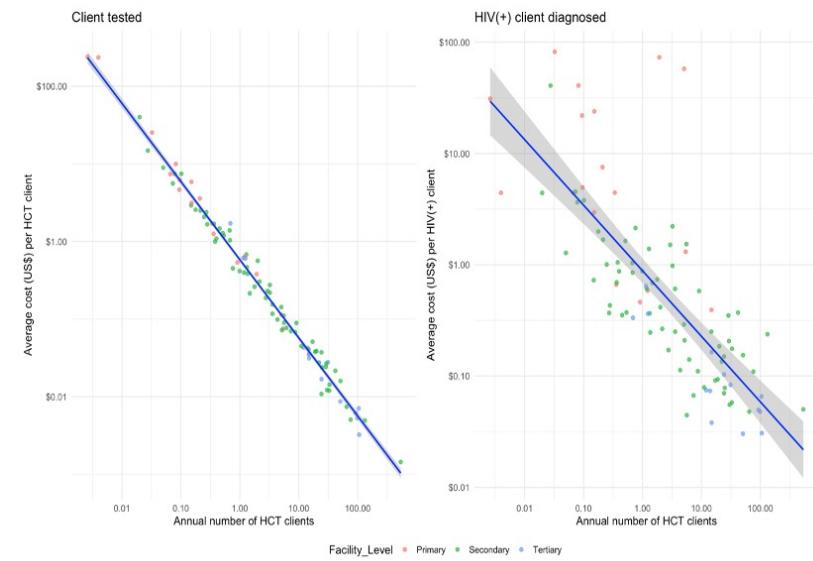
```

1025 Figure 5
1026 ````{r}
1027 # Filter explicitly
1028 HCT_filtered <- HCT %>%
1029 filter(is.na(HTC_tested2)) %>%
1030 slice_head(n = 141)
1031
1032 # Define exchange rate (exactly as original paper)
1033 exchange_rate <- 150
1034
1035 # Compute Total Annual Costs explicitly
1036 HCT_filtered <- HCT_filtered %>%
1037 mutate(
1038 Personnel_Cost = (wage_md + wage_hp + wage_rn + wage_hs + wage_in) / exchange_rate,
1039 Recurrent_Cost = (total_extsup + total_localgov + total_nat + total_comm) / exchange_rate,
1040 Capital_Cost = total_manage / exchange_rate,
1041 Training_Cost = HTC_Vignettes / exchange_rate,
1042 Total_Annual_Cost = Personnel_Cost + Recurrent_Cost + Capital_Cost + Training_Cost
1043)
1044
1045 # Compute correct unit costs clearly
1046 HCT_filtered <- HCT_filtered %>%
1047 filter(HTC_tested2 > 0, HTC_pos > 0, Total_Annual_Cost > 0) %>%
1048 mutate(
1049 Cost_per_client_tested = Total_Annual_Cost / HTC_tested2,
1050 Cost_per_positive_client = Total_Annual_Cost / HTC_pos
1051) %>%
1052 filter(Cost_per_client_tested > 0, Cost_per_positive_client > 0)
1053
1054 # IMPORTANT STEP: Create correct 'Facility_Level' column explicitly:
1055 HCT_filtered <- HCT_filtered %>%
1056 mutate(
1057 Facility_Level = case_when(
1058 level3 == 1 ~ "Tertiary",
1059 level2 == 1 ~ "Secondary",
1060 TRUE ~ "Primary"
1061)
1062) %>%
1063 mutate(Facility_Level = factor(Facility_Level, levels = c("Primary", "Secondary", "Tertiary")))
1064
1065 # Panel 1: Cost per client tested (explicitly matching Fig 5)
1066 plot_tested <- ggplot(HCT_filtered, aes(x = HTC_tested, y = Cost_per_client_tested, color = Facility_Level)) +
1067 geom_point(alpha = 0.7, size = 2) +
1068 geom_smooth(method = "lm", se = TRUE, color = blue) +
1069 scale_x_log10(labels = comma_format()) +
1070 scale_y_log10(labels = dollar_format(prefix = "$")) +
1071 labs(
1072 x = "Annual number of HCT clients",
1073 y = "Average cost (US$) per HCT client",
1074 title = "Client tested"
1075) +
1076 theme_minimal(base_size = 14) +
1077 theme(legend.position = "bottom")
1078
1079 # Panel 2: Cost per HIV(+) client diagnosed (explicitly matching Fig 5)
1080 plot_positive <- ggplot(HCT_filtered, aes(x = HTC_tested2, y = Cost_per_positive_client, color = Facility_Level)) +
1081 geom_point(alpha = 0.7, size = 2) +
1082 geom_smooth(method = "lm", se = TRUE, color = blue) +
1083 scale_x_log10(labels = comma_format()) +
1084 scale_y_log10(labels = dollar_format(prefix = "$")) +
1085 labs(
1086 x = "Annual number of HCT clients",
1087 y = "Average cost (US$) per HIV(+) client",
1088 title = "HIV(+) client diagnosed"
1089) +
1090 theme_minimal(base_size = 14) +
1091 theme(legend.position = "bottom")
1092
1093 # Combine plots side-by-side explicitly and clearly
1094 combined_plot <- plot_tested + plot_positive + plot_layout(guides = "collect") & theme(legend.position = "bottom")
1095
1096 # Explicitly display combined plot
1097 combined_plot
1098
1099 ````
```



**Fig 5. Relationship between HCT unit costs and scale, by cascade stage.**

<https://doi.org/10.1371/journal.pone.0201706.g005>

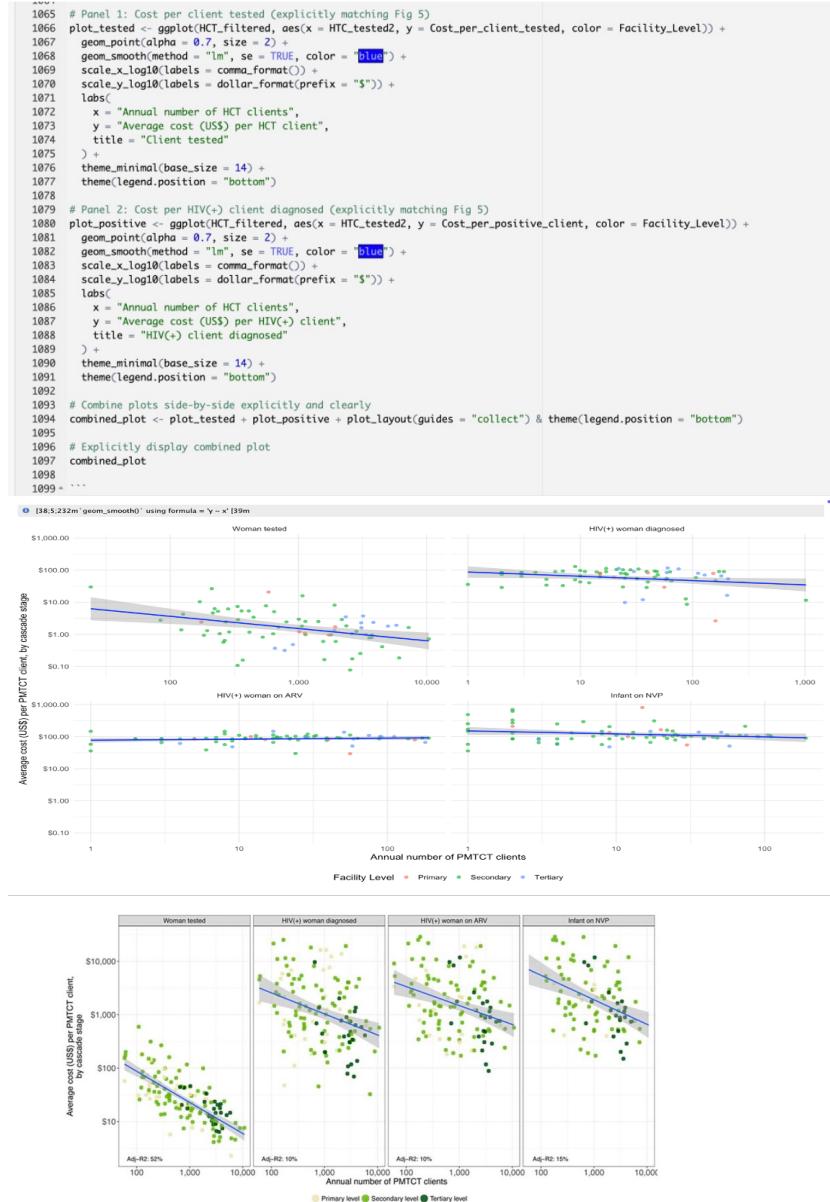


### Description:

- Scatter plots with log-log scales showing the relationship between average cost and service scale. Linear regression lines fitted with 95% confidence intervals. Then separate panels for cost per client tested and cost per HIV(+) client diagnosed. Facility levels shown with different color coding.

### **5.1.9 Replication of Figure 6: PMTCT Unit Costs and Scale Relationship**

The following images show the replication process for Figure 6:



**Fig 6. Relationship between PMTCT unit cost and scale, by cascade stage.**

<https://doi.org/10.1371/journal.pone.0201706.g006>

### Description:

- Scatter plots with log scales showing the relationship between average PMTCT cost and client scale. Separate panels for: woman tested, HIV(+) woman diagnosed, HIV(+) woman on ARV, and infant on NVP. Regres-

sion lines fitted with 95% confidence intervals. Facility levels indicated by color coding (Primary, Secondary, Tertiary).

### 5.1.10 Replication of Figure 7: HCT Regression Models

The following images show the replication process for Figure 7:

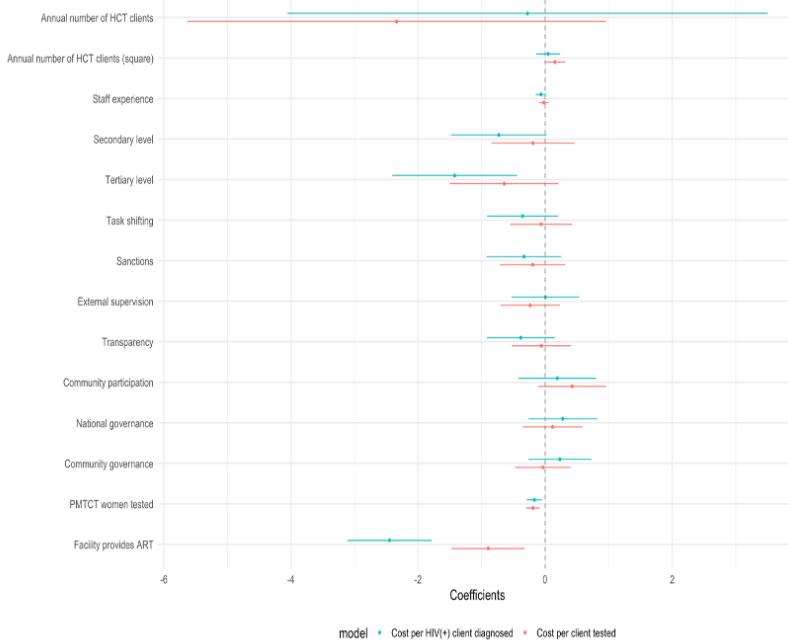
```
1200 Figure 7
1201
1202 + ``{r}
1203 # Create squared log scale variable
1204 HCT_reg <- HCT_filtered %>%
1205 filter(
1206 !is.na(ln_cost_tested),
1207 !is.na(ln_cost_positive),
1208 !is.na(ln_tot_htc)
1209) %>%
1210 mutate(
1211 ln_tot_htc_sq = ln_tot_htc^2
1212)
1213
1214 # Run both models
1215 model_tested <- lm(ln_cost_tested ~ ln_tot_htc + ln_tot_htc_sq + yr_exp_staff +
1216 level2 + level3 + task_shift_htc +
1217 total_sanctyp + total_extsup +
1218 total_transparency2 + total_comm +
1219 total_nat + total_localgov +
1220 PMTCT_test + provide_art,
1221 data = HCT_reg)
1222
1223 model_positive <- lm(ln_cost_positive ~ ln_tot_htc + ln_tot_htc_sq + yr_exp_staff +
1224 level2 + level3 + task_shift_htc +
1225 total_sanctyp + total_extsup +
1226 total_transparency2 + total_comm +
1227 total_nat + total_localgov +
1228 PMTCT_test + provide_art,
1229 data = HCT_reg)
1230
1231 # Cleaned label mapping for display
1232 nice_names <- c(
1233 "ln_tot_htc" = "Annual number of HCT clients",
1234 "ln_tot_htc_sq" = "Annual number of HCT clients (square)",
1235 "yr_exp_staff" = "Staff experience",
1236 "level2" = "Secondary level",
1237 "level3" = "Tertiary level",
1238 "task_shift_htc" = "Task shifting",
1239 "total_transparency2" = "Transparency"
1240)
326:24 [C] Chunk 12 ▾
```

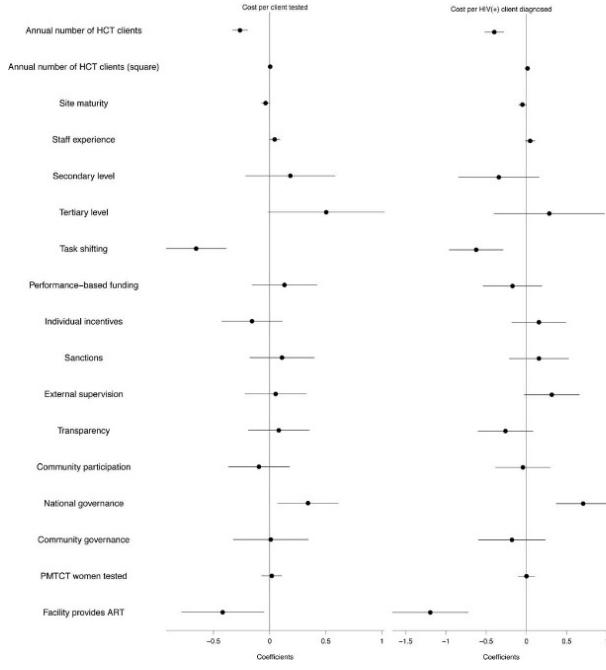
```

1260 total_transparency + total_comm +
1267 total_nat + total_localgov +
1268 PMTCT_test + provide_art,
1269 data = HCT_reg)
1270
1271 # Cleaned label mapping for display
1272 nice_names <- c(
1273 "ln_tot_htc" = "Annual number of HCT clients",
1274 "ln_tot_htc_sq" = "Annual number of HCT clients (square)",
1275 "yr_exp_staff" = "Staff experience",
1276 "level2" = "Secondary level",
1277 "level3" = "Tertiary level",
1278 "task_shift_htc" = "Task shifting",
1279 "total_sanctype" = "Sanctions",
1280 "total_extsup" = "External supervision",
1281 "total_transparency2" = "Transparency",
1282 "total_comm" = "Community participation",
1283 "total_nat" = "National governance",
1284 "total_localgov" = "Community governance",
1285 "PMTCT_test" = "PMTCT women tested",
1286 "provide_art" = "Facility provides ART"
1287)
1288
1289 # Combine models into plot
1290 dwplot(list(
1291 "Cost per client tested" = model_tested,
1292 "Cost per HIV(+) client diagnosed" = model_positive
1293)) %>%
1294 relabel_predictors(nice_names) +
1295 geom_vline(xintercept = 0, linetype = "dashed", color = "gray50") +
1296 theme_minimal(base_size = 14) +
1297 labs(title = "Fig 7. HCT regression models along the service cascade",
1298 x = "Coefficients", y = "") +
1299 theme(legend.position = "bottom")
1300

```

Fig 7. HCT regression models along the service cascade





**Fig 7. HCT regression models along the service cascade.**

<https://doi.org/10.1371/journal.pone.0201706.g007>

#### Description:

- Regression coefficient plots for HCT unit cost models. Separate models for cost per client tested and cost per HIV(+) client diagnosed. Visualized scale measures (number of clients), facility characteristics (level, staff experience), and governance indicators.

#### 5.1.11 Replication of Figure 8: PMTCT Regression Models along the Service Cascade

The following images show the replication process for Figure 8:

```

1263 Figure 8
1264 ````{r}
1265 # Step 1: Prepare your dataset with precise filtering
1266 PMTCT_reg <- PMTCT_filtered %>%
1267 filter(
1268 PMTCT_tested > 0,
1269 PMTCT_positive > 0,
1270 PMTCT_ART_clients > 0,
1271 PMTCT_nvp > 0,
1272 Total_Annual_Cost > 0
1273) %>%
1274 mutate(
1275 ln_tot_pmtct = log(PMTCT_tested),
1276 ln_tot_pmtct_sq = ln_tot_pmtct^2,
1277 Cost_per_woman_tested = Total_Annual_Cost / PMTCT_tested,
1278 Cost_per_HIV_pos_woman = Total_Annual_Cost / PMTCT_positive,
1279 Cost_per_HIV_woman_ARV = Total_Annual_Cost / PMTCT_ART_clients,
1280 Cost_per_infant_NVP = Total_Annual_Cost / PMTCT_nvp
1281) %>%
1282 filter(
1283 Cost_per_woman_tested > 0,
1284 Cost_per_HIV_pos_woman > 0,
1285 Cost_per_HIV_woman_ARV > 0,
1286 Cost_per_infant_NVP > 0
1287)
1288
1289 # Step 2: Run the regressions
1290 model1 <- lm(log(Cost_per_woman_tested) ~ ln_tot_pmtct + ln_tot_pmtct_sq +
1291 yr_exp_staff + level2 + level3 + task_shift_pmtct +
1292 total_sanctyp + total_extsup + total_transparency2 +
1293 total_comm + total_nat + total_localgov +
1294 provide_art,
1295 data = PMTCT_reg)
1296
1297 model2 <- lm(log(Cost_per_HIV_pos_woman) ~ ln_tot_pmtct + ln_tot_pmtct_sq +
1298 yr_exp_staff + level2 + level3 + task_shift_pmtct +
1299 total_sanctyp + total_extsup + total_transparency2 +
1300 total_comm + total_nat + total_localgov +
1301 provide_art,
1302 data = PMTCT_reg)

```

```

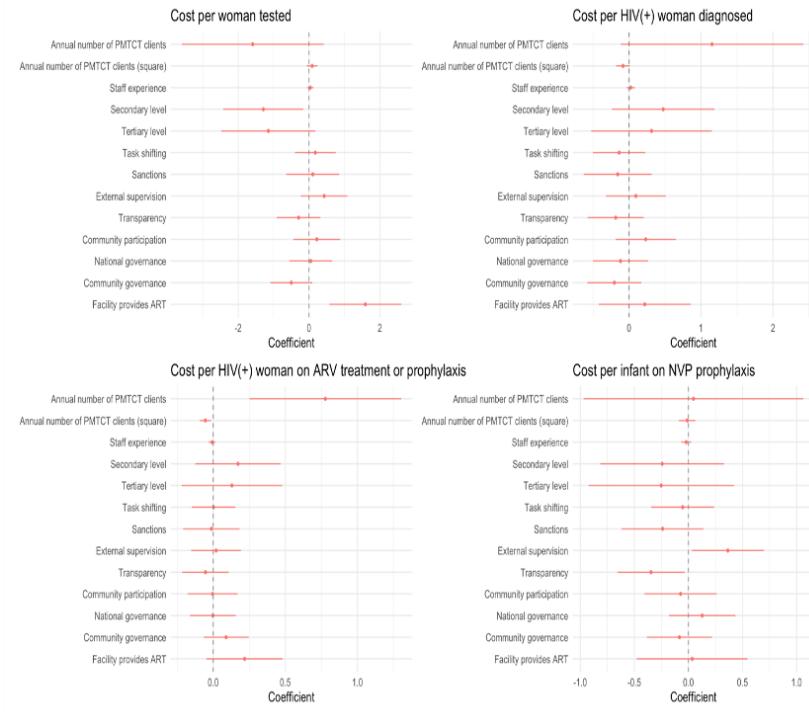
Source Visual
1287
1288
1289 # Step 2: Run the regressions
1290 model1 <- lm(log(Cost_per_woman_tested) ~ ln_tot_pmtct + ln_tot_pmtct_sq +
1291 yr_exp_staff + level2 + level3 + task_shift_pmtct +
1292 total_sanctype + total_extsup + total_transparency2 +
1293 total_comm + total_nat + total_localgov +
1294 provide_art,
1295 data = PMTCT_reg)
1296
1297 model2 <- lm(log(Cost_per_HIV_pos_woman) ~ ln_tot_pmtct + ln_tot_pmtct_sq +
1298 yr_exp_staff + level2 + level3 + task_shift_pmtct +
1299 total_sanctype + total_extsup + total_transparency2 +
1300 total_comm + total_nat + total_localgov +
1301 provide_art,
1302 data = PMTCT_reg)
1303
1304 model3 <- lm(log(Cost_per_HIV_woman_ARV) ~ ln_tot_pmtct + ln_tot_pmtct_sq +
1305 yr_exp_staff + level2 + level3 + task_shift_pmtct +
1306 total_sanctype + total_extsup + total_transparency2 +
1307 total_comm + total_nat + total_localgov +
1308 provide_art,
1309 data = PMTCT_reg)
1310
1311 model4 <- lm(log(Cost_per_infant_NVP) ~ ln_tot_pmtct + ln_tot_pmtct_sq +
1312 yr_exp_staff + level2 + level3 + task_shift_pmtct +
1313 total_sanctype + total_extsup + total_transparency2 +
1314 total_comm + total_nat + total_localgov +
1315 provide_art,
1316 data = PMTCT_reg)
1317
1318 # Step 3: Rename variables for clearer labels
1319 nice_names <- c(
1320 "ln_tot_pmtct" = "Annual number of PMTCT clients",
1321 "ln_tot_pmtct_sq" = "Annual number of PMTCT clients (square)",
1322 "yr_exp_staff" = "Staff experience",
1323 "level2" = "Secondary level",
1324 "level3" = "Tertiary level",
1325 "task_shift_pmtct" = "Task shifting",

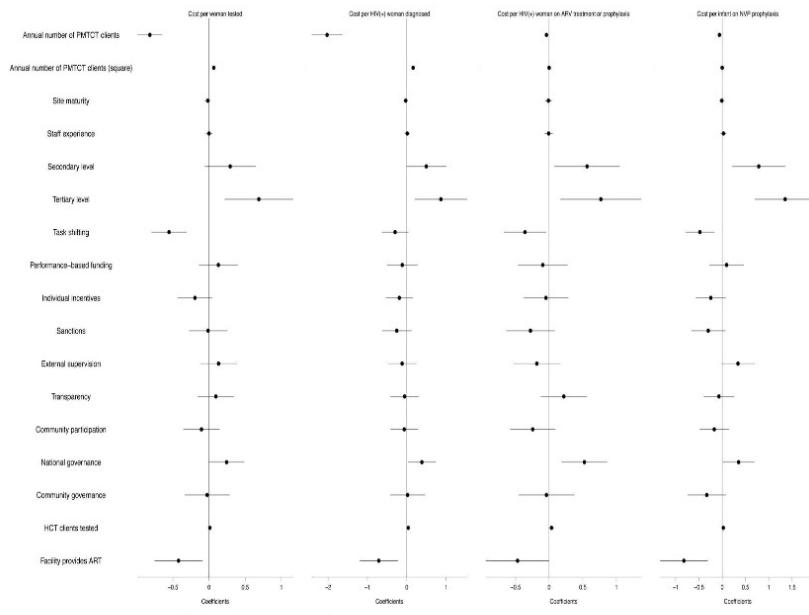
```

```

1318 # Step 3: Rename variables for clearer labels
1319 nice_names <- c(
1320 "ln_tot_pmtct" = "Annual number of PMTCT clients",
1321 "ln_tot_pmtct_sq" = "Annual number of PMTCT clients (square)",
1322 "yr_exp_staff" = "Staff experience",
1323 "level2" = "Secondary level",
1324 "level3" = "Tertiary level",
1325 "task_shift_pmtct" = "Task shifting",
1326 "total_sanctype" = "Sanctions",
1327 "total_extsup" = "External supervision",
1328 "total_transparency2" = "Transparency",
1329 "total_comm" = "Community participation",
1330 "total_nat" = "National governance",
1331 "total_localgov" = "Community governance",
1332 "provide_art" = "Facility provides ART"
1333)
1334
1335 # Step 4: Function to generate individual dot-whisker plot
1336 make_dwplot <- function(model, title) {
1337 dwplot(model) %>%
1338 relabel_predictors(nice_names) +
1339 geom_vline(xintercept = 0, linetype = "dashed", color = "gray50") +
1340 theme_minimal(base_size = 13) +
1341 labs(title = title, x = "Coefficient", y = NULL) +
1342 theme(legend.position = "none")
1343 }
1344
1345 # Step 5: Create each plot
1346 p1 <- make_dwplot(model1, "Cost per woman tested")
1347 p2 <- make_dwplot(model2, "Cost per HIV(+) woman diagnosed")
1348 p3 <- make_dwplot(model3, "Cost per HIV(+) woman on ARV treatment or prophylaxis")
1349 p4 <- make_dwplot(model4, "Cost per infant on NVP prophylaxis")
1350
1351 # Step 6: Arrange all four in a 2x2 grid (like Figure 8 in the journal)
1352 (p1 | p2) / (p3 | p4)
1353
1354 ...

```





**Fig 8. PMTCT regression models along the service cascade.**

<https://doi.org/10.1371/journal.pone.0201706.g008>

### Description:

- Regression models separately run for four PMTCT cost outcomes. Each model controls for annual number of PMTCT clients (and its square), site characteristics, and governance factors. Plots display coefficients with 95 percent of confidence intervals.