Evaluating the Efficacy of Stratified Sampling with Proportional Allocation for Estimation of Trachoma Prevalence in Sudan

By

Sana Charania

B.A. Emory University, 2016

Faculty Thesis Advisor: Juan S. Leon, PH.D., M.P.H

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
Global Epidemiology

2018

ABSTRACT

Evaluating the Efficacy of Stratified Sampling with Proportional Allocation for Estimation of Trachoma Prevalence in Sudan

By Sana Charania

Trachoma is the leading infectious cause of blindness in the world. The Carter Center implemented the SAFE strategy to reduce the prevalence of trachoma in Sudan, which incorporated surveillance activities. Sampling methods used in this effort - Probability Proportional to Estimated Size (PPES) - led to large variability in the sampling weights resulting in imprecise estimates. The objective of this study is to obtain a valid and precise estimate of the prevalence of Trachomatous Trichiasis (TT) among individuals ≥15 in Sudan and inform programming efforts for The Carter Center's trachoma surveillance system. We drew a sample of four districts in Sudan using stratified sampling with proportional allocation. We calculated prevalence estimates and accounted for stratification and clustering when evaluating parameter estimates and their standard errors for each locality. Then, we compared the estimator's qualities of the new design to those of the PPES study to assess the efficacy of the two sampling protocols and determine whether the stratified random sampling approach added precision to the estimator. The prevalence of TT exceeded the WHO threshold of 0.2% in all four localities: Al Kamleen 1.73% (0.95, 3.14), Al Managle 0.81 (0.45, 1.46), Rife Kassala 0.50 (0.20, 1.29) and Sinnar 1.29% (0.79, 2.11). The average overall Design effect (*Deff*) was lower in PPES sampling compared to stratified sampling with Proportional Allocation (1.26). Due to the little variability of Trachoma between clusters, the *Deff* yields a number that indicates the sample was more efficient than a simple random sample (*Deff*<1). This is likely due to small prevalence of the outcome rather than efficacy of the sample design. The *Deff* from the weights were comparable between the two sampling protocols. Evaluating the efficacy of the protocol with a more variable outcome, we calculated the *Deff* for the prevalence of females. Where there is more variability between clusters, the PPES study yields a larger *Deff* (2.56) than stratified sampling (1.90). Extremely low prevalence can make assessing sampling designs difficult, but there is evidence that stratified sampling provides benefits where prevalence of the condition is small and strong evidence of improvement where the prevalence is large.

Evaluating the Efficacy of Stratified Sampling with Proportional Allocation for Estimation of Trachoma Prevalence in Sudan

By

Sana Charania

B.A., Emory University, 2016

Faculty Thesis Advisor: Juan S. Leon, PH.D, M.P.H

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health in Global Epidemiology

2018

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Special thanks to my thesis advisor and mentor, Paul Weiss, for his guidance, expertise and unwavering support not only for my thesis, but also during my time at Rollins. Thank you to Juan Leon for taking the time to read my thesis, provide valuable feedback and ensure the epidemiology department requirements were met. Lastly, thank you to The Carter Center for the opportunity to work on a topic that is of great interest to me.

TABLE OF ABBREVIATIONS

Abbreviation	Description
CV	Coefficient of Variation
Deff	Design Effect
EPSEM	Equal Probability of Selection for Each Member
PPES	Probability Proportional to Estimated SIZE
PPS	Probability Proportional to Size
PSU	Primary Sampling Unit
SUDAAN	Survey Data Analysis
SAFE	Surgery, Antibiotics, Facial Cleanliness, Environmental Factors
SAS	Statistical Analysis System
SRS	Simple Random Sample
TF	Trachomatous inflammation-follicular
TS	Trachomatous Scarring
TT	Trachomatous Trichiasis
WHO	World Health Organization

CHAPTER I: LITERATURE REVIEW

The following literature review focuses on the current state of trachoma surveillance in Sudan and complex sampling methods employed to estimate the prevalence of Trachomatous Trachiasis.

Trachoma Surveillance

Trachoma is a leading infectious cause of blindness in the world [8]. A majority (85%) of all trachoma cases (18 million) are on the African continent. Figure 1 is a graphic designed by The Carter Center depicting the life cycle of trachoma [16]. Trachoma is caused by the bacterium *Chlamydia trachomatis* and easily spread through "direct personal contact, shared towels and cloths, and flies that have come in contact with the eyes or nose of an infected person" [8]. Because of trachoma infection, an estimated 2.2 million people are visually impaired. Of these 2.2 million people, 12 million are blind [9].

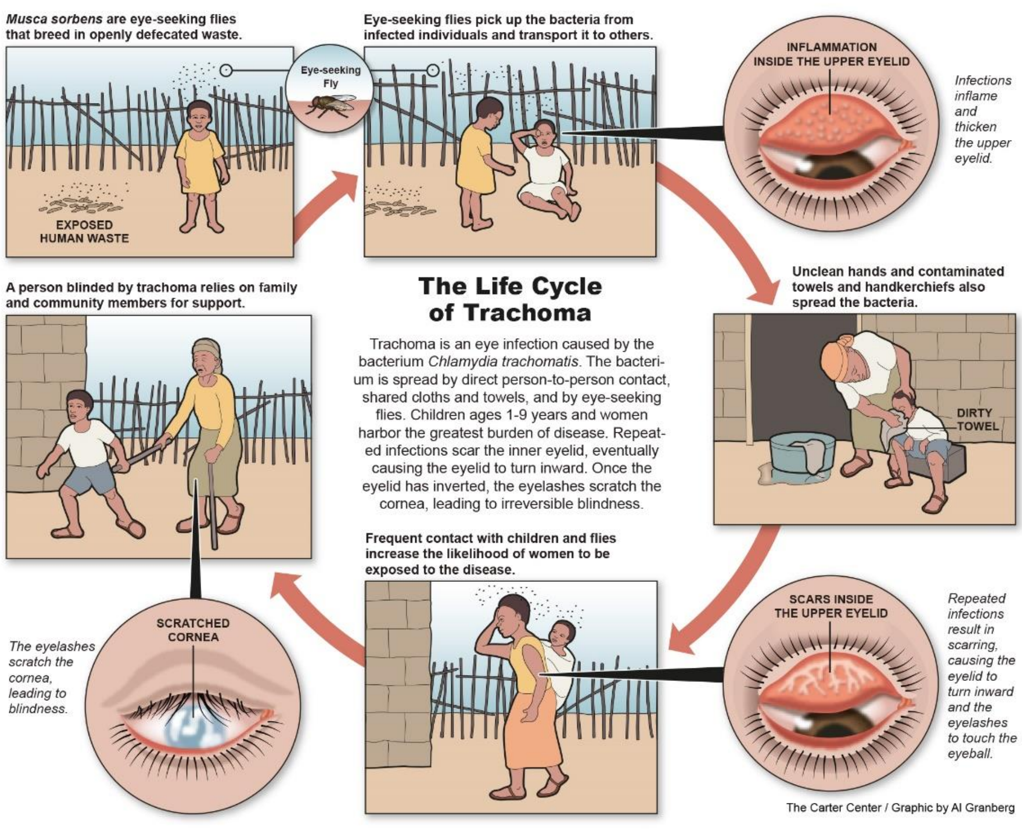


Figure 1. The Life Cycle of Trachoma

The World Health Organization (WHO) simplified grading scale categorizes trachoma into five clinical stages [9]. The first stage, Trachomatous inflammation-follicular (TF), shows the presence of five or more follicles in the upper part of the upper eyelid (tarsal conjunctiva) when the eyelid is inverted. This stage mostly requires topical treatment. Trachomatous inflammation-intense is the second stage of trachoma and has pronounced inflammation of the tarsal conjunctiva. For this stage, topical and systemic treatments are used. The third stage, Trachomatous Scarring (TS) is when the tarsal conjunctiva has visible scars, which may "obscure tarsal blood vessels" [9]. Repeated infections and scarring can cause the eyelid to turn inward. This results in the fourth stage, Trachomatous Trichiasis (TT), a painful state where the eyelashes rub against the cornea. This stage is when an individual is referred for eyelid surgery. TT can cause long-term damage the cornea, leading to vision impairment or blindness. Permeant blindness due to trachoma can lead to the fifth stage, Corneal Opacity [9].

The WHO has set the goal to eliminate trachoma by the year 2020 [10]. WHO defines the elimination of trachoma as the reduction of prevalence of TT "'unknown to the health system' to less than 1 case per 1000 total population ('known' cases are those in whom TT has recurred after surgery, those who refuse surgery, or those yet to undergo surgery whose surgical date is set); and a reduction in the prevalence of the active trachoma sign 'TF' in children aged 1–9 years to less than 5%" [10]. In addition to the disability resulting from the disease, there is a high economic burden of trachoma [10]. The cost on affected communities and individuals is between "US$ 2.9 - 5 billion annually, increasing to US$ 8 billion when trichiasis is included" [10]. However, surgery costing roughly US$ 40 per surgery and "a large donation of azithromycin, preventive chemotherapy" is a cost-effective, affordable way to prevent overspending [10].

As a result, The Carter Center implemented a strategy to decrease the prevalence of trachoma in Sudan, previously one of the countries with the highest prevalence of trachoma, and has seen tremendous progress. The SAFE strategy is an "innovative, multifaceted public health strategy" curated by the WHO [8].

- (S) surgery to correct advance stages of TT
- (A) mass drug administration of antibiotics such as azithromycin
- (F) facial cleanliness and
- (E) environmental improvement such as improved sanitation and hygiene [8]

The Carter Center has implemented the SAFE strategy in endemic localities in Sudan to decrease the prevalence of trachoma (The Carter Center, personal communication, April 18, 2018). The Carter Center has used two types of surveys to assess the prevalence of trachoma: Baseline and impact surveys are surveys aimed at estimating the prevalence of TF among children ages 1-9 [11] in order to determine if a locality is endemic for trachoma and whether the implementation of the SAFE strategy is warranted. All consenting individuals in a household are examined. As a result, an artifact of baseline/ impact surveys is the prevalence of TT among adults 15 years and older. The second type of survey is TT only. TT surveys are not routine. The following is an excerpt from the publication "Design and validation of a trachomatous trichiasis-only survey" describing under what circumstances TT only surveys may be warranted.

> 1) If at baseline survey, the estimated prevalence of TF in 1–9-year-olds is < 5% and of TT in adults is ≥ 0.2%, an impact survey to again measure the TF prevalence is

not indicated; after interventions, a TT-only survey to re-estimate the TT prevalence is indicated.

2) If at surveillance survey, the estimated prevalence of TF in 1–9-year-olds is < 5% and of TT in adults is ≥ 0.2%, further surveys to again measure the TF prevalence are not indicated; after interventions, a TT-only survey to re-estimate the TT prevalence is indicated.

3) If a survey at any stage of the programme estimated the prevalence of TT with a questionable methodological approach, the programme may wish to conduct a TT-only survey.

4) If at baseline survey, the estimated prevalence of TF in 1–9-year-olds is ≥ 30% and of TT in adults is ≥ 0.2%, at least 5 years of A, F and E interventions are recommended before an impact survey to again measure the TF prevalence. During this time, the programme may wish to undertake a TT-only survey to assess progress in addressing the TT backlog, facilitating adjustments in delivery of S interventions, if needed. [15]

Additionally, only consenting adults 15 years or older are examined in TT only surveys.

The Carter Center's efforts to reduce the prevalence of trachoma in Sudan led to an Impact survey aided by the WHO employing Probability Proportional to Estimated Size (PPES) to estimate and assess the prevalence of trachoma in Sudan (The Carter Center, personal communication, October 17, 2017). There were issues with the PPES sampling design that led to large increases in variance estimates [14]. A major issue with the PPES design was improper estimation of cluster (village) sizes in Sudan. Large deviation of the estimated cluster size from true cluster size led to a non-EPSEM outcome and great variability in the sampling weights.

Surveys and Post-Validation Surveillance are all required under the WHO guidelines [11]. However, surveys involving "special issues" or "any specific efforts to investigate trachoma prevalence and/or interventions in difficult to reach populations" or studying special circumstances that may have affected the program are optional [11].

The Carter Center is reassessing the prevalence of trachoma in various districts of Sudan using an alternative sampling approach to the PPES design. We hypothesize that stratified sampling with proportional allocation should improve the quality of the survey and will inform future TT reduction efforts Sudan. Due to the severity of TT, appropriate management of individuals with TT is a priority of all trachoma elimination programs. Obtaining valid and precise data on TT prevalence is imperative to allow programs to plan, implement, monitor and evaluate interventions.

Complex Designs:

Simple Random Sample (SRS) designs are considered the gold standard of sampling methods and provide unbiased, representative samples and estimates [1]. However, a SRS is rarely feasible with large populations. There are four main advantages to sampling methods compared to taking a census: reduced cost, greater speed, greater scope and greater accuracy [1]. When done correctly, a sample can yield results that are more accurate due to higher quality of information in comparison to a census [1]. The following section will cover advantages and disadvantages of key methods used to estimate the prevalence of TT in Sudan.

Equal Probability of Selection for Each Member:

Each element in an Equal Probability of Selection for Each Member (EPSEM) design has the same probability of selection. This is achieved when each element in the sample is given the same weight, allowing the weights to cancel out and leaving the estimator unchanged [1]. The following proof demonstrates how weights cancel out under EPSEM conditions.

$$\pi_i = \Pr\ (i^{th} \text{ element is selected into the sample})$$

$$W_i = \frac{1}{\pi_i}$$

$$\bar{y}_w = \frac{\Sigma\, w_i\, y_i}{\Sigma\, w_i}$$

Under EPSEM conditions, all weights are the same:

$$\bar{y}_w = \frac{\Sigma\, w\, y_i}{\Sigma\, w}$$

$$= \frac{w\, \Sigma\ y_i}{n\, w}$$

$$= \frac{\Sigma\ y_i}{n}$$

$$= \bar{y}$$

Stratification:

Strata are "non-overlapping, homogenous groupings of population elements" [2]. The sampling statistician forms strata "prior to the selection of probability sample" [2]. For example, populations can be stratified by race or gender. Stratification produces a gain in precision of the estimates when you can divide a heterogeneous population into sub populations that are homogenous [1]. Under ideal conditions, stratification can lead to large gains in precision. Increasing the samples precision is achieved by forming "strata that are 'homogeneous within' and 'heterogeneous between'" [2]. This is achieved under three circumstances. Firstly, when the population is composed of groups varying largely in size. Secondly, when the variables measured are closely related to the size of the group. Lastly, a good measure of size is available for creating the strata [1] It is important to note that an improvement in the precision of the estimator is not seen when the strata is independent of the outcome.

Clustering:

Clusters are naturally occurring groups in a population [1]. Neighborhoods and schools are both examples of clusters due to natural homogeneity in the unit. Dissimilarly from stratification, clustering generally decreases precision in estimates. An increase in design effect is caused by the correlations within the clusters. Many characteristics measured on sample elements within a cluster are correlated. For example, elements like socioeconomic status, political attitudes and access to facilities are all characteristics that clusters may share to a greater or lesser degree. As a result, "the amount of 'statistical information' contained in a clustered sample of *n* persons is less than in an independently selected simple random sample of the same size" [2]. Therefore, clustered sampling generally leads to increases in standard errors and design effects. [2].

Design Effect:

The Design Effect (*Deff*) is defined as "The ratio of the variance of the estimate obtained from the (more complex) sample to the variance of the estimate obtained from the simple random sample of the same number of units" [1]

$$\textit{Sampling Deff} = \frac{Var\ (\theta)_{cs}}{Var(\theta)_{srs}}$$

The *Deff* allows us to determine the survey sample size and evaluate the efficiency of more complex sampling designs compared to a SRS of the same size [3]. A design effect coefficient of one indicates that the sampling design is equivalent to a SRS in efficiency/precision. An increase in the *Deff* to greater than one indicates some loss of precision in the estimate compared to a SRS. One sampling method that is known to cause a loss of precision in the estimates is clustering. Lastly, a design effect smaller than one indicates the sampling design shows increased precision in relation to a SRS, one such complex design is stratification [1]. Furthermore, the design effect from the weights are calculated by assessing the variability in the weights over the mean squared of the weights, or one plus the Coefficient of Variation (CV) squared. This tells us how much variability in the design is due to the weights [4].

$$\textit{Deff from Weights} = 1 + \frac{Var(Weights)}{Mean\ (Weights)^2} = 1 + CV^2$$

The Overall *Deff* incorporates both the *Deff* from the weights and the *Deff* from the sample.

$$\text{Overall } \textit{Deff} = (\textit{Deff}_{w} \times \textit{Deff}_{sample})$$

Probability Proportional to Estimated Size:

Probability Proportional to Estimated Size (PPES) sampling is used in lieu of Probability Proportional to Size (PPS), “a finite population in which a size measure is available for each population unit before sampling” [5]. PPES is employed when the exact cluster size is unknown.

PPES allows larger Primary Sampling Unit (PSU) a greater probability of being selected [5]. Smaller PSUs are less likely to be selected, but weigh more in the analysis. If the sampling is done correctly and the primary unit sizes are correct, the design should yield a close to EPSEM sample with unequal-sized clusters. Huge differences between larger and smaller PSUs as well as the sample size leads to non-constant weights and more variability. The variability in PPES multi-stage designs depends on the primary units, and the degrees of freedom are calculated based on the number of PSUs, not the final sample size. The probabilities will be more unstable if the size estimates differ greatly from the truth. Deviation from the truth can result in flawed designs. Additionally, selection probabilities no longer cancel in the PPES design as they do PPS sampling.

Proportional Allocation:

Proportional Allocation is a method used to select clusters from each stratum. In this method, the sampling statistician draws the same proportion of people from each stratum [6]. Thus, the "number of units selected from each stratum directly depends on the number of units in the stratum" [6]. Additionally, Proportional Allocation "provides a self-weighting sample; estimates can be made with greater speed and a higher degree of precision" [6]. One major advantage of sampling this way is that it is a simple method to use. However, any deviation from the design can lead to a non-EPSEM sample.

The flawed PPES design led to large variability in the sampling weights which resulted in imprecise point estimates for the prevalence of trachoma. The objective of this thesis is to obtain valid and precise estimates of Trachoma in Sudan and to inform future efforts of The Carter Center's trachoma surveillance system. Understanding and correctly employing these complex design methods is vital to obtaining precise estimates, reducing cost of survey's and ultimately reducing the burden of blindness from trachoma.

CHAPTER II: MANUSCRIPT

Introduction:

Trachoma is a leading infectious cause of blindness in the world [8]. A majority (85%) of all trachoma cases (18 million) are on the African continent. Trachoma is caused by the bacterium *Chlamydia trachomatis* and easily spread through "direct personal contact, shared towels and cloths, and flies that have come in contact with the eyes or nose of an infected person" [8]. As a result of trachoma infection, an estimated 2.2 million people are visually impaired. Of these 2.2 million people, 1.2 million are blind [9].

The WHO has set the goal to eliminate trachoma by the year 2020. WHO defines the elimination of trachoma as the reduction of prevalence of TT "'unknown to the health system' to less than 1 case per 1000 total population ('known' cases are those in whom TT has recurred after surgery, those who refuse surgery, or those yet to undergo surgery whose surgical date is set); and a reduction in the prevalence of the active trachoma sign 'TF' in children aged 1–9 years to less than 5%" [10]. As a result, The Carter Center implemented the SAFE strategy to decrease the prevalence of Trachoma in Sudan, previously one of the countries with the highest prevalence of Trachoma, and has seen tremendous progress. The SAFE strategy is an "innovative, multifaceted public health strategy" curated by the WHO [10].

- (S) surgery to correct advance stages of TT
- (A) mass drug administration of antibiotics such as azithromycin
- (F) facial cleanliness and
- (E) environmental improvement such as improved sanitation and hygiene [10]

These efforts led to a study aided by the WHO using a Probability Proportional to Estimated Size (PPES) design to estimate and assess the prevalence of trachoma in Sudan (The Carter Center, personal communication, October 17, 2018). There were issues with the PPES sampling design that led to fairly large increases in variance estimates (The Carter Center, personal communication, October 17, 2018). A major issue with the PPES design was improper estimation of cluster (village) sizes in Sudan. Large deviation of the estimated cluster size from the true cluster size leads to a non-EPSEM outcome and great variability in the sampling weights. The Carter Center is reassessing the prevalence of trachoma in various districts of Sudan using an alternative sampling approach to the PPES design. We hypothesize that stratified sampling with proportional allocation should improve the quality of the survey and will inform future TT reduction efforts Sudan. Due to the severity of TT, appropriate management of individuals with TT is a priority of all trachoma elimination programs. Obtaining valid and precise data on TT prevalence is imperative to allow programs to plan, implement, monitor and evaluate interventions.

Methods

Ethical Statement

Emory University determined that IRB approval was not needed for these surveys as they did not meet the criteria for "research" or "clinical investigation" (IRB# 076-2006) No personal identifiers were collected during these surveys. The Informed Consent was explained to participants in the local language and verbal consent was obtained from individuals willing to participate. Follow-ups were included to ensure that all TT cases received treatment at either a medical camp or local clinic. Surgeries were scheduled in conjunction with the trachoma program activities (The Carter Center, personal communication, October 17, 2017).

Study Site and study population

The Carter Center has implemented the SAFE strategy in endemic localities in Sudan to decrease the prevalence of trachoma (The Carter Center, personal communication, April 18, 2018). Subsequently, population based cross-sectional surveys were conducted in six localities in Sudan. The impact survey study population consisted of all consenting individuals and children. The TT only survey population consisted of individuals 15 years of age or older. Table 1 provides descriptive statistics of the population in each locality including percent male and female, average age, total households surveyed and the number of people examined.

Probability Proportional to Estimated Size

The sample size for the PPES study was estimated at each locality assuming a 95% confidence interval and an expected TF prevalence of 3%. A design effect of 3.0, precision of 2% and 20% non-response rate yields a minimum sample size of 1,004 children per locality. Researchers assumed there were 4.7 individuals per household and children ages 1-9 made up 35% of the household. This required them to sample a total of 611 households. The minimum samples size of 1,004 children was reached.

Estimating the prevalence of TT in adults 15 years and older was a secondary research question (The Carter Center, personal communication, October 17, 2017). A total of 5,344 consenting individuals over the age of 15 were surveyed.

Villages were selected from an enumerated list using a systematic random approach, which employed PPES sampling. To achieve the desired sample size of TF, researchers sampled a total of 25 clusters per locality and 25 households per cluster. All households were grouped into five-household segments, then five segments in each cluster were randomly selected and all households in the segment was included. All consenting members present in the household were examined (The Carter Center, personal communication, October 17, 2017). The methods and rational behind all sampling decisions for the PPES survey were mandated by the WHO.

Stratified Sampling with Proportional Allocation

The sample size was estimated at each locality assuming a 95% confidence interval, absolute precision and an expected TT prevalence of 0.2%. A design effect of 1.47 and 10% non-response rate yields a minimum sample of 3,100 number of adults 15 years or older targeted for examination. To achieve the desired sample size and based on the number of households that could be visited in one day within a cluster, we sampled a total of 30 clusters per locality and 35 households in each cluster. The WHO determined these cutoffs. The sampling method was a 2-stage design, which included the selection of clusters, segments and households. In the first stage, localities were stratified into high population villages and low population villages. The cutoff for a high population village was at least 2,000 people. This number was derived from previous sampling frames in Sudan. Then, 30 villages were randomly selected proportional to the percentage of villages in each stratum. The households were selected using segmentation sampling. All households were grouped into five-household segments. Seven segments in each cluster were randomly selected and all households in the segment were included. All consenting individuals 15 years and above were examined from each household (The Carter Center, personal communication, October 17, 2017). We surveyed 6,704 number of adults 15 years or older.

Statistical Analysis

We evaluated two different sampling protocols and quantified the strengths of the stratified sampling with proportional allocation over the previous method from the PPES study. We calculated TT prevalence estimates and confidence intervals using SAS-callable

SUDAAN statistical software [13]. We accounted for the clusters, strata and weights and calculated the estimates with replacement and without replacement to compare how accounting for the sampling design affects the CIs and sample design effects. TT Prevalence estimates were calculated for each locality. Stratification and clustering were accounted for when estimating parameter and variance as available. The estimates of the new design were compared to those of the previous PPES study to measure the efficacy of the sampling protocol and determine whether the new, stratified sampling approach with proportional allocation, improved the estimator of TT. We assessed the relative precision of the estimates when controlling for the design of the study and evaluated the relative impact of the weights derived from each design by accounting for the weights' design effect. This was calculated by the design effect formula elaborated in Kish et al 1987 below for each locality [4].

$$\textit{Deff from Weights} = 1 + \frac{Variance(Weights)}{Mean\ (Weights)^2} = 1 + CV^2$$

We calculated the overall design effect for each locality by multiplying the sample *Deff* provided by SUDAAN to the *Deff* from the weights [4]. Then, we took the average overall *Deff* for both sampling methods.

$$\textit{Overall Deff} = (Deff_w \times Deff_{sample})$$

Results

Table 1 describes the survey sample of both study types. The percent females surveyed was higher in all localities across both sampling types: Abu Jebaiha 1,876 (54.88), Baladayat el Gedarif 1,675 (52.28), Al Kamleen 1,476 (54.28), Al Mangale 1, 338 (55.82), Rife Kassala 1,049 (56.55) and Sinnar 1,454 (53.14). The average age of participants in the impact survey between both localities was 17.56. The average age of participants in the TT only survey among all four localities was 38.08.

Table 2 summarizes the prevalence of TT and design effect from the PPES study, with and without replacement. The precision of the estimator increases as we account for sampling without replacement. PPES had much larger variability in the weights compared to Proportional Allocation (Table 3). TT prevalence is above the WHO threshold of 0.2% in Baladyat el Gedarif: 0.23 (0.79, 2.11). In table 2, we can see that the largest variability in the weights for PPES is the locality Abu Jebaiha, with a variance of 599,450.55. The Average overall design effect is 1.26.

Table 3 summarizes the prevalence of TT and design effect from stratified sampling with proportional allocation, with and without replacement. The precision of the estimator increases as we account for sampling without replacement. TT prevalence was above WHO guidelines for elimination of trachoma of 0.2% in all four localities: Al Kamleen 1.73% (0.95, 3.14), Al Managle 0.81 (0.45, 1.46), Rife Kassala 0.50 (0.20, 1.21) and Sinnar 1.29% (0.79, 2.11) (Table 3). The locality with largest variance in weights in table 3, sampled with proportional allocation, was Al Managle (6,174.23). The average overall design effect is 2.14.

Both localities in the PPES sample had lower design effects and smaller CIs (Table 2). However, all localities from both sampling methods had comparable design effects from the weights (Table 2 and Table 3).

Table 4 summarizes the design effect with the variable sex for each sampling method. In this comparison, PPES yields larger average overall design effects compared to stratified sampling with proportional allocation. The Average Overall *Deff* for PPES is 2.56 and the Average Overall *Deff* for stratified sampling with proportional allocation is 1.90.

Figure 1 depicts the variability in the weights of the PPES design. The mean is represented by the diamond symbol marker and the median is represented by the horizontal line in the box. The mean and median of the weights do not align within Abu Jebaiha or

Baladyat el Gedarif, thus suggesting there is not symmetry in the weights. The final weights range from 112.67 to 3,110.58.

Figure 2 portrays the variability in the weights of stratified sampling with proportional allocation. The mean (represented by the diamond symbol marker) and median (represented by the line inside the box) of the weights in three out of four localities align in the boxplot (Al Kamleen, Rife Kassala and Sinnar), thus suggesting symmetry in the weights. The final weights range from 3.66 to 282.86.

Discussion

We calculated valid and precise estimates of TT prevalence in four localities in Sudan using the stratified sampling approach. Our results indicate that the *Deff* and CIs from the PPES sampling design was smaller than a stratified sampling plan with proportional allocation. The smaller CIs however are due to the larger sample size of the study. Because, the PPES study was designed to detect TF prevalence, which is larger than TT prevalence, the confidence intervals for TT would be narrower than a study specifically designed to detect TT prevalence (The Carter Center, personal communication, October 17, 2017). To evaluate the efficacy of the protocols we looked at the *Deff* of each study.

The average overall *Deff* (1.26) for the PPES study was due to the fact the TT in the PPES sample was an extremely rare condition (table 2). In the first locality, Abu Jebaiha, the prevalence was 0.03% (1 case in the 2,701 individuals examined). The prevalence in the second locality, Baladyat el Gedarif, the prevalence was 0.23% (7 cases in the 2,654 individuals examined) (Table 2). Due to the little variability of Trachoma between the clusters, the sample design effect yields numbers that indicates the sample was more efficient than a simple random sample *(Deff<1):* Abu Jebaiha (Sample *Deff* = 0.82*)* and Baladyat el Gedarif (Sample *Deff* = 0.94*)*. To evaluate the efficacy of the protocol with a more variable outcome, we calculated the *Deff* for the prevalence of sex. Table 4 indicates the results of the sample *Deff*, the variance in the design effects due to the weights and the overall *Deff*. When there is more variability between the clusters, the PPES study yields a larger overall *Deff* (2.56) than the stratified approach (1.90) and the sample *Deff* in the PPES design no longer indicate that the sample was more efficient than a SRS (*Deff<1)*: Abu Jebaiha (Sample *Deff*= 1.62) and Baladyat el Gedarif (Sample *Deff*=1.95) (Table 4).

We also found that the PPES study yielded smaller CIs due to the larger sample size and rare outcome. In a previous study employing PPES sampling "The Prevalence of Blinding Trachoma in Northern States of Sudan" the confidence intervals vary greatly [14]. Furthermore, the confidence intervals for TF, a more prevalent condition are much wider than the confidence intervals of TT [14]. This suggests that if the prevalence of TT was higher, then the CI would be wider and more unstable and wider as well. PPES sampling yields smaller design effects and CIs when the outcome is rare and the sample size is large, but we do not see the same effects in outcomes that are more prevalent and have more variability between clusters (Table 4).

In both cases, we were limited in our analysis because an important variable to account for the design was unavailable. In both the PPES and Proportional Allocation sampling methods, researchers split the clusters (villages) into segments then randomly selected a number of segments to sample households from. To fully account for the design of the study, the segment information must be entered into the NEST and TOTCNT statements in SUDAAN statistical software [13]. However, the data that indicated what households were in each segment was not collected. Thus, the results from our analysis does not take into account the full design. This likely yields smaller CIs and design effects in both sampling methods than would have otherwise been calculated.

Moving forward, we recommend The Carter Center collect all segment information to fully account for the design of the study. We were unable to conduct a cost analysis comparing each sampling method due to lack of data. Obtaining valid and precise data on TT prevalence is imperative to allow programs to plan, implement, monitor and evaluate interventions on strict budgets. One of the main benefits of sampling is reducing costs [1]. A future cost analysis would be beneficial to determine the overall cost of each sample design, the cost per cluster and the cost per case found. We hypothesize that the stratified approach has the greater benefit of reducing cost and increasing accuracy when the outcome is more prevalent.

TABLES

Table 1. *Descriptive Statistics of the Sample*

	Probability Proportional to Estimated Size (Impact Survey)		Proportional Allocation (TT only)			
	Abu Jebaiha	Baladyat el Gedarif	Al Kamleen	Al Mangale	Rife Kassala	Sinnar
Total Males n (%)	1535 (45.12)	1529 (47.72)	1243 (45.72)	1,059 (44.18)	806 (43.45)	1,282 (46.86)
Total Females n (%)	1876 (54.88)	1675 (52.28)	1,476 (54.28)	1,338 (55.82)	1,049 (56.55)	1,454 (53.14)
Age Mean (std)	20.52 (17.55)	18.46 (17.52)	37.53 (17.67)	38.27 (18.17)	38.41 (16.50)	38.24 (17.79)
Average Age for Impact Survey	19.52 (17.56)		**Average Age for TT only**	38.08 (17.61)		
Total Households Surveyed	604	610	849	899	826	880
Examined n (%)	2,701 (84.30)	2,654 (78.01)	1,832 (67.38)	1,670 (69.73)	1,386 (74.44)	1,816 (66.01)

Table 2. *Trachomatous Trichiasis (TT) in those ≥15 years, Probability Proportional to Estimated Size (Impact Survey)*

				With replacement		Without replacement			
Locality	**Examined n (%)**	**TT cases, n**	**Variance in Weights**	TT% (lower and upper bounds)	Sample *Deff*	TT% (lower and upper bounds)	Sample *Deff*	*Deff* from Weights	Overall *Deff*
Abu Jebaiha	2,701 (84.30)	1	599,450.55	0.03 (0.00, 0.25)	0.85	0.03 (0.20, 1.21)	0.82	1.56	1.27
Baladyat el Gedarif	2,654 (78.01)	7	87,756.02	0.23 (0.10, 0.51)	0.90	0.23 (0.79, 2.11)	0.94	1.33	1.25

Average Overall *Deff* 1.26

Table 3. *Trachomatous Trichiasis (TT) in those ≥15 years, Proportional Allocation (TT only)*

Locality	Examined n (%)	TT cases, n	Variance in Weights	With replacement		Without replacement			
				TT% (lower and upper bounds)	Sample *Deff*	TT% (lower and upper bounds)	Sample *Deff*	*Deff* from Weights	Overall *Deff*
Al Kamleen	1,832 (67.38)	29	394.44	1.73 (0.92, 3.23)	3.01	1.73 (0.95, 3.14)	2.47	1.42	3.51
Al Managle	1,670 (69.73)	14	6,174.23	0.81 (0.46, 1.42)	1.03	0.81 (0.45, 1.46)	1.13	1.56	1.76
Rife Kassala	1,386 (74.44)	7	46.74	0.50 (0.21, 1.24)	1.31	0.50 (0.20, 1.21)	1.20	1.36	1.63
Sinnar	1,816 (66.01)	21	132.97	1.29 (0.78, 2.13)	1.39	1.29 (0.79, 2.11)	1.33	1.26	1.68

Average Overall *Deff* 2.14

Table 4. *Evaluating the Design Effect with the variable Sex, without replacement*

	Probability Proportional to Estimated Size (Impact Survey)		Proportional Allocation (TT only)			
	Abu Jebaiha	Baladyat el Gedarif	Al Kamleen	Al Managle	Rife Kassala	Sinnar
Percent Female (Lower and upper bounds)	53.07 (50.71, 55.41)	54.29 (51.82, 56.74)	52.85 (50.66, 55.03)	56.30 (53.39, 59.18)	56.75 (54.69, 58.78)	53.37 (51.14, 55.59)
Sample *Deff*	1.62	1.95	1.25	1.94	0.76	1.30
***Deff* from Weights**	1.56	1.33	1.52	1.56	1.36	1.26
Overall *Deff*	2.52	2.59	1.90	3.03	1.04	1.64
Average Overall *Deff*	**2.56**		**1.90**			

FIGURES

Figure 1. *Variability in Weights, Probability Proportional to Estimated Size (Impact Survey).* The whiskers are drawn from the quartiles to the extreme values of the group. The 75th percentile is the upper edge of the box and the 25th percentile is the lower edge of the box. The median 50th percentile is represented by the line inside the box. The diamond symbol marker represents the mean of the sampling weights. The minimum value is represented by the endpoint of the lower whisker. Circular symbol markers outside of the whiskers represent outliers in the sampling weights.

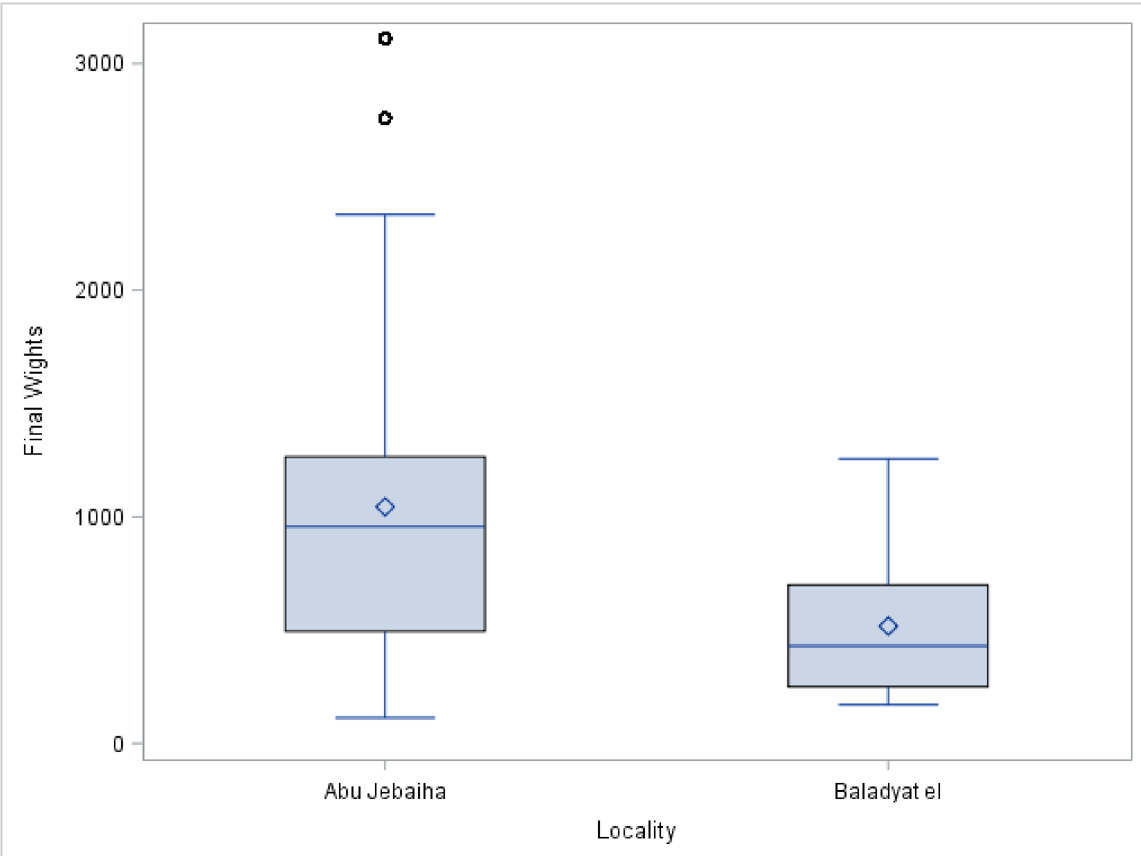
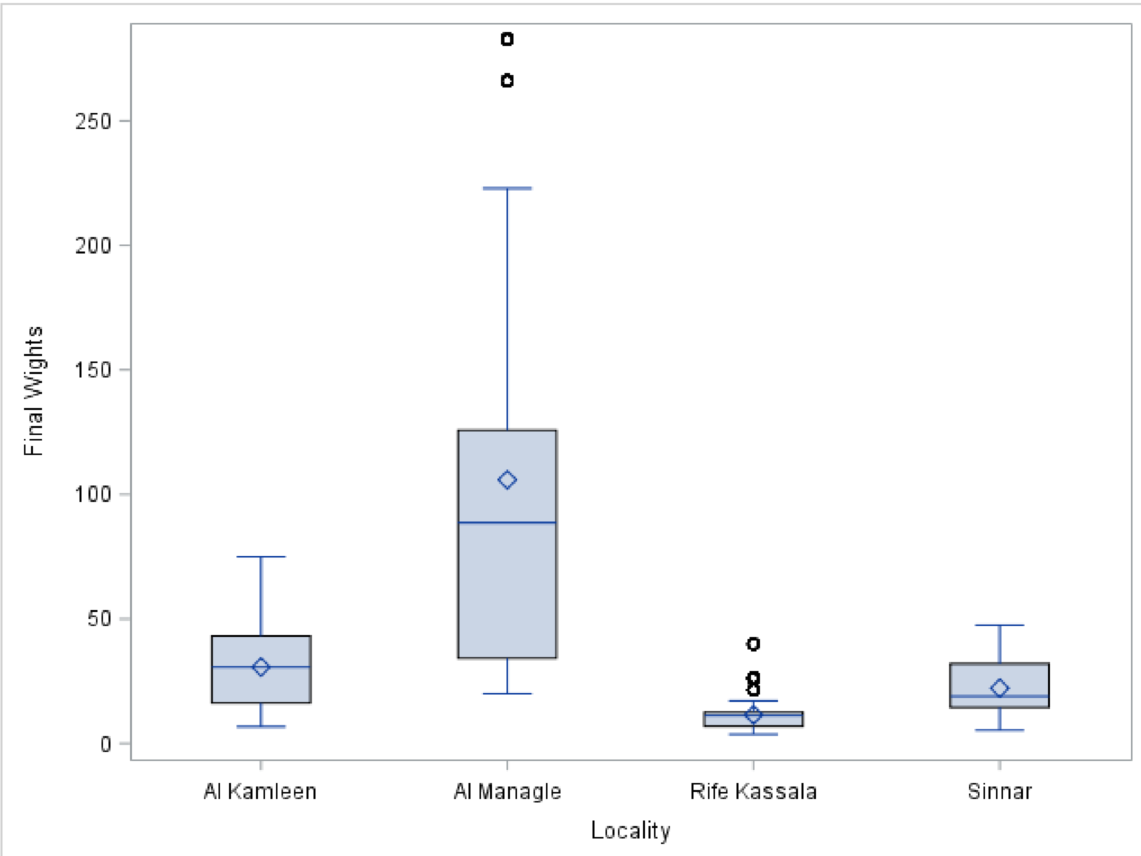


Figure 2. *Variability in Weights, Proportional Allocation (TT only).*
The whiskers are drawn from the quartiles to the extreme values of the group. The 75th percentile is the upper edge of the box and the 25th percentile is the lower edge of the box. The median 50th percentile is represented by the line inside the box. The diamond symbol marker represents the mean of the sampling weights. The minimum value is represented by the endpoint of the lower whisker. Circular symbol markers outside of the whiskers represent outliers in the sampling weights.



REFERENCES

1. Cochrin WG. *Sampling Techniques*. 3rd ed. Oxford, ENG: John Wiley & Sons, Publishers; 1977.
2. Heeringa G, Brady TW, Patricia AB. *Applied Survey Data Analysis*. 1st ed. Boca Raton, FL Chapman and Hall/CRC, Publishers; 2010.
3. Kish L. *Survey Sampling*. 1st ed. Oxford, ENG: John Wiley & Sons, Publishers; 1965
4. Kish, L.*Weighting in Deft2*. The SurveyStatistician. The Newsletter of the International Association of Survey Statisticians, 1987.
5. Skinner C,. Probability Proportional to Size (PPS) Sampling. *Wiley Online Library*. 2016
6. World Health Organization. *Steps in applying Probability Proportional to Size (PPS) and calculating Basic Probability Weights,*. Geneva, CH
7. Pandey R. and Verma M.R. *Samples Allocation in Different Strata for Impact Evaluation of Developmental Programme*. Rev. Mat. Estat., 26 (4), 103–112., 2008.
8. World Health Organization. *Progress report on elimination of trachoma, 2013*, Weekly epidemiological record. 2014, Geneva, Switzerland. p. 421-428.
9. Centers for Disease Control and Prevention. *Hygiene-Related Diseases*, Centers for Disease Control and Prevention. 2009, Atlanta,GA.
10. World Health Organization. *Global elimination of blinding trachoma*, *51st World Health Assembly*. 1998, World Health Assembly: Geneva.
11. World Health Organization. *Trachoma*. 2018, World Health Organization: Johns Hopkins University. 2010, Baltimore, MA.
12. World Health Organization. *Validation of Elimination of Trachoma as a Public Health Problem,* World Health Organization, Geneva. 2016

13. RTI International. SUDAAN Statistical Software, release 11.0.1 Research Triangle Park, NC: RTI International; 2018.

14. Hassan A, Ngondi J, King J, et al. The Prevalence of Blinding Trachoma in Northern States of Sudan. *PLOS Neglected Tropical Diseases.* 2011.

15. World Health Organization. *Design and Validation of a trachomatous trachiasis-only survey,* World Health Organization, Geneva. 2017

16. The Carter Center. *The Life Cycle of Trachoma.* The Task Force for Global Health, 2016.

CHAPTER III: SUMMARY, PUBLIC HEALTH IMPLICATIONS, POSSIBLE FUTURE DIRECTIONS

The Carter Center implemented the SAFE strategy to reduce the prevalence of trachoma in Sudan, which incorporated surveillance activities (The Carter Center, personal communication, October 17, 2017). Sampling methods used in this effort - Probability Proportional to Estimated Size - led to large variability in the sampling weights and variance (The Carter Center, personal communication, April 18, 2018). The objective of this study was to obtain a valid and precise estimate of the prevalence of Trachomatous Trichiasis (TT) among individuals 15 years of age and older in Sudan and inform programming efforts for The Carter Center's Trachoma surveillance system. A sample of four districts in Sudan were drawn using a stratified sampling plan with Proportional Allocation. Prevalence estimates were calculated for the four localities to evaluate the effectiveness of SAFE on prevalence of TT in the region. We evaluated the estimator's qualities of the new design to those of the previous PPES study to assess the efficacy of the two sampling protocols and determine whether the new stratified random sampling approach added precision to the estimator. We met our objective and obtained valid and precise estimates of TT prevalence in Sudan. The prevalence of TT exceeded the WHO threshold of 0.2% in all four localities Al Kamleen 1.73% (0.95, 3.14), Al Managle 0.81 (0.45, 1.46), Rife Kassala 0.50 (0.20, 1.21) and Sinnar 1.29% (0.79, 2.11). The design effect among samples was lower in PPES sampling: Abu Jebaiha (0.85) and Baladyat el Gedarif (0.90). This was likely due to small prevalence rather than efficacy of the sample design. The *Deff* from the weights were comparable between the two sampling protocols. Due to the little variability of Trachoma between clusters, the design effect yields a number that indicates the sample was more efficient than a simple random sample. Evaluating the efficacy of the protocol with a more variable outcome, we calculated the design effect for the prevalence of females. Where there is more variability between clusters, the PPES study yields a larger *Deff* than stratified sampling. Extremely low prevalence can make assessing sampling designs difficult, but there is evidence that stratified sampling provides benefits where prevalence of the condition is small and strong evidence of improvement where prevalence is large.

We were limited to our analysis because an important variable to account for the design was unavailable. In both survey types, researchers split the clusters (villages) into segments then randomly selected a number of segments to sample households from. The segment information must be entered into the NEST and TOTCNT statements in SUDAAN statistical software. However, the data that indicated what households were in each segment was not collected. Thus, the results from our analysis does not take into account the full design. Moving forward, we recommend The Carter Center collect all segment information to fully account for the design of the study. This will likely affect the way The Carter Center collects data and accounts for the design in subsequent studies, regardless of which sampling method they use.

Furthermore, PPES sampling method and the variables collected to analyze the data is a fairly common practice among trachoma surveillance. This study will improve trachoma surveillance by advising surveillance systems to employ an alternative, more robust, method. Our findings suggest that when an outcome is more prevalent, stratified sampling with proportional allocation yields better estimates and smaller design effects than PPES. Additionally, we found that it is fairly common in surveillance practice to not collect which households are in each segment (The Carter Center, personal communication, April 18, 2018). We recommended that The Carter Center and other trachoma surveillance systems improve their survey design and analysis specifically by collecting all segment data to in order to fully account for their study design. Thus, overall our study results will likely impact the way The Carter Center proceeds to conduct TT only and Impact Surveys moving forward.

We were unable to conduct a cost analysis comparing each sampling method due to lack of data. As previously discussed, one of the main advantages of sampling is reduced cost of the design. We hypothesize stratified sampling with proportional allocation to assess the prevalence of trachoma is more cost-effective than PPES. A future cost analysis would be beneficial to determine the overall cost of each sample design, the cost per cluster and the cost per case found. Moving forward, we will conduct a cost analysis of each sampling method when the data are available from Sudan. Obtaining valid and precise data on TT prevalence is imperative to allow programs to plan, implement, monitor and evaluate interventions on strict budgets.