10.1071/SH12056_AC

Supplementary Material: Sexual Health 10(1), 95-96.

Validation of fixed size sampling plans using resampling: specifications

Let $N_w$denote the sample size demanded by the sampling plan (w = 20, 30, 40, 50, 60, 70 or 80 warts) and let X_k be the kth simulated data set (k= 1...100),representing 100 sampled warts with the proportion p_k being negative for HPV 6/11. One hundred distinct values of prevalence,p_k,were sampled using Latin hypercube sampling (a powerful sampling method to generate a representative set of parameter estimates)(1,2)from a uniform distribution covering the range (0.00-0.19) of non-vaccine HPV prevalence reported in the literature.(3–8) For validation of sampling performance, 1000 resampling iterations were carried out on each of these 100 data sets and performance assessed by summary statistics. A resampling iterationconsisted of taking 100 samples of size N_w from each data set X_k, using sampling with replacement. Sampling with replacement is computationally efficient and produces results that differ little from sampling without replacement in numerical validation of sampling plans.(9)

Assessing sampling precision

Precision (D) was defined as

$$D = \frac{s_k}{\sqrt{N_w}} \Big/ \bar{x}_k \quad (1)$$

In equation (1), and $\bar{x}_k$ are the standard deviation and mean, respectively, of the number of warts negative for HPV 6/11 over a resampling iteration from the k^{th} data set that used samples

of size N_w. Because D is based on sample estimates, D itself takes on different values for different samples, so over 100 different data sets there is some variability in the achieved D for different sample sizes – see Figure 1. The main reason a sample size of 60 is better than a sample size of 50 is because the median and mean are closer together (less skewed distribution of D), and because no improvement in this skewness is observed with a sample size of greater than 60 (see Figure 1 in this Supplementary Material).

Assessing the probability of making an incorrect classification using the plans

The operating characteristic (OC) curve was fitted to assess the probability that using a particular sample size would indicate that the proportion of warts negative for HPV 6/11 has not changed from pre-vaccination programme levels, designated P_T. The type I and II error rates in Table 1 in the main text are an alternative way of representing the OC function. The OC curves themselves are shown in Figure 2 in this Supplementary Material. The estimate of P_T used was 10%, equal to the mean of the uniform distribution fitted to the published data (see Methods in the main paper). The observed OC values for each data set X_k were plotted against the proportion of warts in the data set negative for HPV 6/11, and a four parameter logistic OC curve was fitted to the scatter. The OC curve was described by the function

$$OC = y_0 + \frac{a}{1+\left(\frac{P_{LR}}{x_0}\right)^b} . \quad (2)$$

In equation (2), y_0 and a are lower and upper limits of the logistic function (set at 0 and 1), P_{LR} is the proportion of resampling iterations where the proportion of the sample negative for HPV

6/11 was greater than P_T, and x_0 is the value of P_{LR} at the point of inflection of the curve. The exponent *b* is a slope parameter. The curve was fitted using SigmaPlot 8.0 (Systat, Richmond, CA, USA).

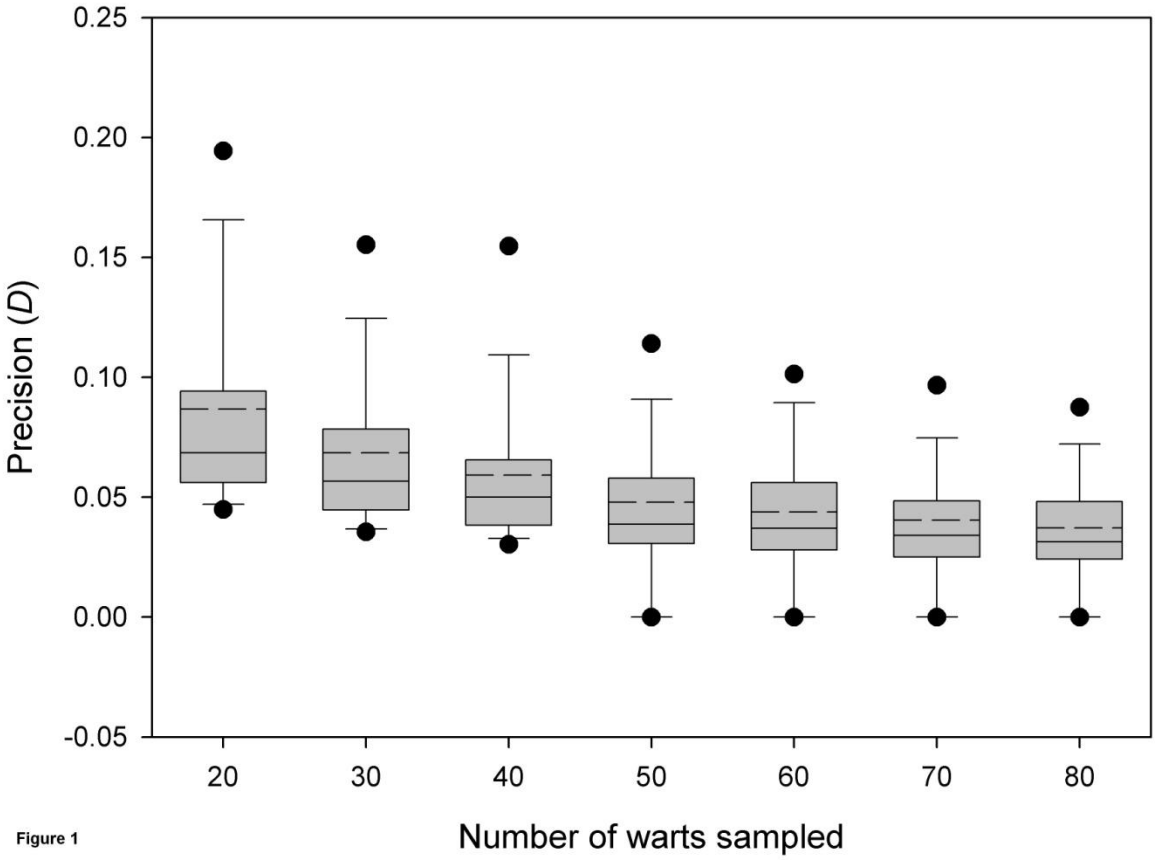

Figure 1 Achieved precision (*D*) of fixed sample size plans when implemented using resampling on 100 data sets with different prevalence. Decreasing values of *D* represent increasing confidence in the sample estimate of the proportion negative for HPV 6/11.The solid and dotted lines in each box represent the median and mean precision achieved, respectively; the bars show the 10th and 90th percentiles, and the circles are the 5th and 95th percentiles. Variability in *D* over

data sets of different prevalence decreases with increasing sample size, with the difference between mean and median *D* being least with a sample size of > 60 warts.

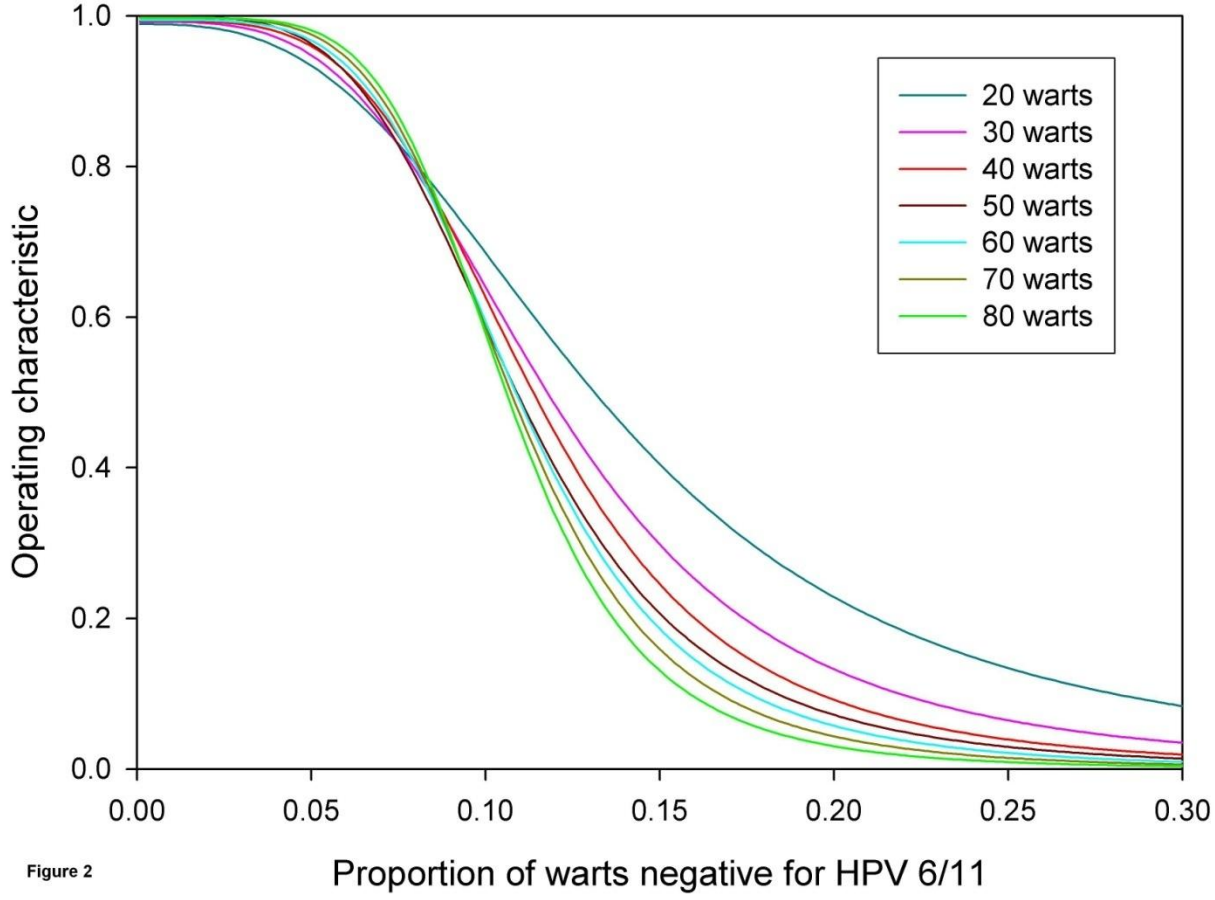

Figure 2 Operating characteristic (OC) curves fitted to the proportion of warts negative for HPV 6/11 per simulated data set using each fixed sample size (key in legend). The OC describes the probability that sampling will indicate that the proportion of warts negative for HPV 6/11 is < 0.10. Not surprisingly the probability of an incorrect decision in this regard is highest when

prevalence is 0.10 and decreases to low levels with reasonable sample size (> 60) as prevalence increases.

References

1. Hamilton AJ, Waters EK, Kim HJ, Pak WS, Furlong MJ. Validation of Fixed Sample Size Plans for Monitoring Lepidopteran Pests of Brassica oleracea Crops in North Korea. J Econ Entomol. 2009 Jun;102:1336–46.

2. Crowley PH. Resampling methods for computation-intensive data analysis in ecology and evolution. Annual review of ecology and systematics [Internet]. 1992 [cited 2011 May 30];23:405–47. Available from: http://cat.inist.fr/?aModele=afficheN&cpsidt=4527460

3. Garland SM, Steben M, Sings HL, James M, Lu S, Railkar R, et al. Natural History of Genital Warts: Analysis of the Placebo Arm of 2 Randomized Phase III Trials of a Quadrivalent Human Papillomavirus (Types 6, 11, 16, and 18) Vaccine. J Infect Dis. 2009;199(6):805–14.

4. Brown DR, Schroeder JM, Bryan JT, Stoler MH, Fife KH. Detection of Multiple Human Papillomavirus Types in Condylomata Acuminata Lesions from Otherwise Healthy and Immunosuppressed Patients. J Clin Microbiol. 1999;37:3316–22.

5. Grce M, Husnjak K, Skerlev M, Lipozencic J, Pavelic K. Detection and typing of human papillomaviruses by means of polymerase chain reaction and fragment lenth polymorphism in male genital lesions. Anticancer Res. 2000;20:2097–102.

6. Müller EE, Chirwa TF, Lewis DA. Human papillomavirus (HPV) infection in heterosexual South African men attending sexual health services: associations between HPV and HIV serostatus. Sex Transm Infect. 2010 Jun 1;86:175–80.

7. Potocnik M, Kocjan B, Seme K, Poljak M. Distribution of human papillomavirus (HPV) genotypes in genital warts from males in Slovenia. Acta Dermatovenerol Alp Panonica Adriat. 2007;16:91–6.

8. Chan PKS, Luk ACS, Luk TNM, Lee K-F, Cheung JLK, Ho K-M, et al. Distribution of human papillomavirus types in anogenital warts of men. J Clin Virol. 2009 Feb;44:111–4.

9. Naranjo SE, Hutchison WD. Validation of Arthropod Sampling Plans Using a Resampling Approach: Software and Analysis. Am Entomol. 1997;43:48–57.