Published in final edited form as:

Am J Obstet Gynecol. 2013 February ; 208(2): 112.e1–112.e11. doi:10.1016/j.ajog.2012.11.040.

Development and comparison of prognostic scoring systems for surgical closure of genitourinary fistula

Vera FRAJZYNGIER, PhD, MPH, Guohua LI, MD, DrPH, Elaine LARSON, PhD, RN, FAAN, CIC, Joseph RUMINJO, MD, and Mark A. BARONE, DVM, MS

EngenderHealth, New York, NY (VF, JR, MAB), the Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY (VF, GL, EL), the Department of Anesthesiology, College of Physicians and Surgeons, Columbia University, New York, NY (GL) and the School of Nursing, Columbia University, New York, NY (EL)

Abstract

Objective—To test the diagnostic performance of five existing classification systems (developed by Lawson, Tafesse, Goh, the World Health Organization (WHO) and Waaldijk), and a prognostic scoring system derived empirically from our data, to predict fistula closure three months following surgery.

Study Design—Women with genitourinary fistula (n=1274) presenting for surgical repair services at 11 health facilities in sub-Saharan Africa and Asia were enrolled in a prospective cohort study. Using one-half the sample we created multivariate generalized estimating equation models to obtain weighted prognostic scores for components of each existing classification system, and the empirically-derived scoring system. With the second half, we developed Receiver Operating Characteristic curves using the prognostic scores and calculated areas under the curves (AUCs) and 95% confidence intervals (CI) for each system.

Results—Among existing systems, the scoring systems representing the WHO, Goh and Tafesse classifications had the highest predictive accuracy: AUC 0.63 (95%CI: 0.57-0.68), AUC 0.62 (95%CI: 0.57-0.68) and AUC 0.60 (95%CI: 0.55-0.65), respectively. The empirically-derived prognostic score achieved similar predictive accuracy (AUC 0.62, 95%CI: 0.56-0.67); it included significant predictors of closure found in the other classification systems, but contained fewer, non-overlapping components. Differences in AUCs were not statistically significant.

Conclusions—The prognostic values of existing urinary fistula classification systems and the empirically-derived score were poor to fair. Further evaluation of the validity and reliability of existing classification systems to predict fistula closure is warranted, with consideration given to a prognostic score that is evidence-based, simple and easy to use.

© 2012 Mosby, Inc. All rights reserved.

Reprint requests: Vera Frajzyngier, PhD, MPH, vfraczyngier@gmail.com. **Corresponding author:** Vera Frajzyngier, PhD, MPH, 308 5th Avenue, Brooklyn, NY 11215; Phone: 347-628-1016; vfraczyngier@gmail.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosure of interests: The authors report no conflict of interest.

Paper presentation: This research was presented as an oral presentation at the 20th International Federation of Obstetrics and Gynecology (FIGO) World Congress of Obstetrics and Gynecology, Rome, Italy, October 7-12 2012.

Condensation: Existing fistula classification systems and an empirically-derived prognostic score have poor to fair ability to predict fistula closure.

Short version of title: Comparison of prognostic scores for surgical genitourinary fistula closure

Keywords

Genitourinary fistula; surgery; classification; observational study; receiver operating characteristics

Introduction

While only garnering worldwide attention in the past decade, female genitourinary fistula, an abnormal opening between the genital and urinary systems, is an ancient condition, predominantly caused by obstructed labor. From the mid-19th century, when the first consistently successful surgical techniques for repairing genitourinary fistulas were developed, efforts have been made to develop a schema for classifying them.¹ At least 25 systems exist,² although the reliability and validity of most have not been empirically tested. While there is widespread acknowledgement that a standardized classification system is needed,²⁻⁶ disagreement remains about which fistula characteristics should be included, and what purposes (e.g. prognostic or descriptive) the system should serve.

The purposes of existing systems and the components they include vary. They are used for didactic purposes, to facilitate communication and learning, and for planning and conducting repairs, including assessing prognosis and determining the need for referral. Some systems, particularly older ones (e.g. Sims,¹ Lawson⁷), describe the location of the fistula only. Others, (e.g. Goh,⁸ Tafesse⁹ and Waaldijk¹⁰) are more detailed, describing the extent to which varying anatomical structures are affected, as well as factors such as bladder and fistula size. The more detailed systems allow for a precise description of the fistula, with the implicit assumption that as type increases by number or letter combination (e.g. Type 2Bb versus Type 2A), the prognosis worsens. Indeed, the systems developed by Goh and Waaldijk have been empirically tested to determine the extent to which their parameters predict repair outcomes.^{11,12} An additional system presented by the World Health Organization (WHO)¹³ classifies fistula on the degree of repair difficulty (simple or complex). However, to our knowledge, this system has not been validated, nor is it currently used. None of the systems we are aware of are scoring systems and none include patient characteristics, including comorbidities.

These systems were developed using clinical judgment, rather than empirical evidence. Few studies have examined the ability of individual patient or fistula characteristics to predict fistula repair outcomes and the evidence-base on most predictor-repair outcome relationships remains thin.¹⁴ One recent study directly compared Goh and Waaldijk's classification systems; while providing an important contribution to the evidence-base, limitations included a small sample and short follow-up length.¹⁵

A standardized evidence-based prognostic classification system would facilitate communication and learning across fistula services and assist with patient triage and selection.⁴ An evidence-based prognostic *scoring* system in particular would have unique advantages. A scoring system could facilitate: (1) surgeons' decisions regarding patient referral, by providing thresholds for what constitutes a "good" or "poor" prognosis; (2) comparison of studies that examine treatment outcomes; (3) evaluation of surgical success rates across facilities and (4) effectiveness of interventions independent of confounding by indication.

To be clinically and analytically useful, a classification system must be both simple and sufficient. A simple and sufficient system would be more readily used and would increase

prognostic accuracy. For analytic purposes, it would need to decrease opportunities for residual confounding, yet not over-adjusting and unnecessarily increasing variance.¹⁶

Using data collected as part of a multi-country prospective cohort study, our primary aim was to test the diagnostic performance of five classification systems (Lawson's,⁷ Waaldijk's,¹⁰ Tafesse's,⁹ Goh's,⁸ and WHO¹³) to predict fistula closure. These systems are either commonly used in clinical settings (Waaldijk and Goh), or represent a range of detail, from more sparse (Lawson) to more exhaustive (Tafesse and WHO). Our secondary aim was to evaluate whether the inclusion of patient or fistula characteristics not included in other classification systems is warranted, or whether existing systems could be simplified, for prognostic purposes.

Methods

Study participants and procedures

1389 women who had surgery for repair of a genitourinary fistula at 11 sites in Bangladesh, Guinea, Niger, Nigeria and Uganda were enrolled between September 2007 and September 2010. All sites were hospitals or clinics receiving support from EngenderHealth's Fistula Care project to conduct repairs. Twenty-five women were excluded because they underwent repair for recto-vaginal fistula only. An additional 35 women were excluded because they were referred to other facilities, did not have surgery for medical/safety reasons, or were treated by catheterization. Excluded women were evenly distributed across all facilities. The majority (95.9%) returned for 3 month follow-up; these 1274 women constitute the sample for these analyses.

Data were collected by site staff on standardized case report forms used at all sites. Site staff carrying out the study were trained in study procedures and interview techniques. Prior to surgery, information was collected on socio-demographics, obstetric history, clinical exam results and medical care provided. At the time of surgery, detailed information was collected on fistula characteristics, intra-operative procedures performed and surgical outcomes. Prior to discharge, data on post-operative care provided and surgical outcomes were recorded. Three months following surgery surgical outcomes were assessed during a clinical exam.

National level ethical approval was obtained in Nigeria, Uganda, Guinea, and Niger. Facility-based ethical review was required and obtained at two of three facilities in Bangladesh; the third gave permission for the study to be conducted. All patients provided signed informed consent (if the patient was not literate, consent was indicated via thumbprint and a witness signed the form).

Measures

The primary outcome was genitourinary fistula closure (versus "not closed") three months following surgery. Closure was assessed via pelvic exam, with a dye test when there was leakage of urine. At two sites, pelvic exams were not routinely conducted. In these cases (186 women; 14.6%), closure was determined using the question "does the client have continuous and uncontrolled leakage of urine?", with a dye test to assess fistula closure in any patient complaining of urine leakage.

We directly measured many components of existing classifications in our data set; however, in some cases it was necessary to create variables closely corresponding to these components using measures in our data set. The operationalization of components of the classification systems using variables from our dataset is detailed in Table 1. There are no agreed upon standard definitions for many fistula characteristics; thus, degree of scarring and tissue loss, and bladder size, were subjectively assessed by operating surgeons.

We also evaluated whether variables not included in existing classification systems merited inclusion in a scoring system, or whether variables already included should be revised or re-categorized. In particular, we evaluated individual characteristics including patient age, fistula duration, and comorbidities prior to surgery. Age and duration of the fistula were measured as continuous variables. Comorbidities assessed included presence of malnutrition (yes versus no, as determined through either a skin fold measurement, body mass index or visual assessment), anemia (yes versus no, as determined through either hemoglobin level, hematocrit or visual assessment), urinary tract infection (UTI, based on clinician reports), and parasitic infections, including malaria (based on clinician reports). Finally, we examined the distributions of ordinal variables included in existing classification systems to determine whether cut-points should be revised.

Statistical analyses

We employed a split-sample design. Half the sample (the derivation cohort) was used to create scoring systems representing the five existing classifications and one scored empirically derived from our data. The second half (the validation cohort) was used to test the scores.

Socio-demographics and repair outcomes in the two cohorts were compared using t-tests for continuous variables and Chi-square tests or Fisher's exact tests (where cell sizes were less than 5) for ordinal or dichotomous variables.

Characteristics of patients whose fistulas were closed at the 3 month follow-up visit were compared to those whose fistulas were not closed using risk ratios (RRs) and corresponding 95% confidence intervals (CIs); RRs were reported instead of odds ratios (ORs) for all analyses because the likelihood of failed fistula closure was greater than 10%, and therefore ORs would overestimate rather than approximate RRs. RRs and 95% CIs were derived using generalized estimating equations (GEE), using an exchangeable correlation structure with a robust standard error estimator to account for clustering of patient outcomes by facility; results accounting for clustering by attending surgeon (not shown) were similar. RRs were generated using the logarithm link function and binomial distribution specification in SAS PROC GENMOD.¹⁷

Using the derived cohort, we constructed separate multivariate GEE models for the components of each classification system. RRs were generated using log-binomial models; in two models where the log-binomial model failed to converge, SAS PROC GENMOD's Poisson regression capability with a log-link function and robust variance was used.¹⁸ Weighted scores for individual classification system components were derived from adjusted RRs (ARRs); scores were only assigned to those components significant at p-value<.05. Weights were rounded to the nearest whole number.

The multivariate model used to develop the empirically-derived score included variables that were associated with repair failure at p-value <.20 in bivariate analysis or were conceptually associated with repair outcome. In the event that two candidate variables were inter-correlated, the variable with the most clinical significance was included.

Using the validation cohort, sensitivity and specificity were calculated for each scoring system. Receiver Operating Characteristic (ROC) curves depicting the relationship between the proportion of true-positives and false-positives for each system were drawn and compared visually, and areas under the ROC curves (AUCs) and 95% CIs were calculated for each curve. Using methods for paired data, AUCs were compared by calculating the contrast chi-square and corresponding p-value for the difference between the AUCs. ROC curves and AUCs were also generated using the derivation cohort to assess model

robustness. ROC curves and AUCs were also generated using the derivation cohort to assess model robustness (Figure 1 and Table 5). All analyses were done using SAS version 9.2; AUCs were calculated using the %roc macro¹⁹ and ROC curves were constructed using the %rocplot macro.²⁰

Results

Baseline characteristics and repair outcomes were similar between the derivation and validation cohorts (Table 2). The proportions of successful fistula closure at three months were 81.5% and 82.0% in the derived and validation cohorts, respectively.

Results of bivariate and multivariate analyses used to develop scores for the existing classification systems are shown in Table 3. Variables independently associated with closure in multivariate analysis received scores based on the ARR. After adjusting for other components of the Goh classification system, urethral length and vaginal scarring or a small bladder independently predicted failure of fistula closure, and the four components of the system representing these variables were scored. Only one component of Lawson's classification, midvaginal location, was significantly associated with repair outcome and therefore scored. Since an ROC curve cannot be created with only one operating point, this system was not analyzed further. The Tafesse system had the most components that were scored. After adjusting for other components, patients with Class 3 and Class 4 fistulas, those with either no or complete involvement of the middle third of the urethra or complete destruction of the urethra, and those with extensive tissue damage or an obliterated vagina were significantly more likely to have a fistula that was not closed than those without these components. While representing a small proportion of the sample, patients with Waaldijk's Type 2 fistulas (specifically Type2Aa and Type2Bb) were at increased risk of having a fistula that was not closed compared to patients without these fistulas, and these two variables were therefore scored. In the model representing WHO's classification system, having greater than one fistula, involvement of the urethra/continence mechanism and extensive tissue loss were all independent predictors of failure of fistula closure, and received scores.

Finally, we empirically derived a prognostic score based on significant predictors of failed closure in other classification systems and factors not included in other classification systems that predicted failure of fistula closure at $p < 0.20$ in bivariate analysis; additional bivariate analyses to inform the multivariate model representing an empirically-derived system are shown in Table 4. Duration of fistula and prior repair were moderately correlated ($r = .36$); we included the latter in our multivariate model because it is a component of existing classification systems and had fewer missing observations in our dataset. Similarly, moderate and extensive tissue loss and moderate and extensive scarring were correlated ($r = .52$); we included moderate or extensive scarring in our final model, as it may be more objectively measured and has been evaluated in prior studies. We excluded closing mechanism involvement: "closing mechanism" may be understood as damage to the urethral sphincter, or to the combination of anatomical structures that contribute to continence.²¹ Thus, some surgeons may not have characterized a woman as having a damaged closing mechanism if the urethral sphincter was intact but other components of the continence mechanism were damaged, leading to underestimation of this measure. Other variables included in the multivariate model for the empirically-derived score were fistula size, the presence of necrotic tissue, cervix not visible, bladder size, and the component of Waaldijk's classification system "ureteric and other exceptional fistulas." Components that were statistically significant after adjusting for other factors were scored based on the ARRs; the final empirically-derived prognostic scoring system contained greater than one fistula (ARR 2.05, 95% CI: 1.28-3.29), moderate or severe scarring (ARR 1.57, 95% CI: 1.12-2.19), partial

urethral involvement (ARR 1.39, 95%CI: 1.05-1.84), and complete destruction of the urethra or transection/circumferential injury (ARR 2.37, 95%CI: 1.80-3.11).

A total score was generated for each participant for each classification system, using the scored components of each system described above, and these results were used to plot ROC curves (Figure 1 and Figure 2) and to calculate corresponding AUCs (Table 5). The WHO, Goh, Tafesse, and WHO systems, and the empirically-derived score had similar ($p=.47$) discriminatory values: AUC 0.63 (95%CI: 0.57-0.68), AUC 0.62 (95%CI: 0.57- 0.68), AUC 0.60 (95%CI: 0.55-0.65), and AUC 0.62 (95%CI: 0.56-0.67), respectively. The Waaldijk classification had a 51% probability of correctly distinguishing patients with failure of fistula closure from those whose fistula were successfully closed (95%CI: 0.49-0.52), significantly lower than other systems.

Comment

The WHO, Goh, and Tafesse classification systems, as well as the empirically-derived prognostic score had similar predictive values for fistula closure at 3 months following repair surgery. However, none had what would be considered good predictive accuracy (AUC > 0.70). The low AUCs suggest that factors other than fistula characteristics, such as surgeon skill (especially critical given the reconstructive nature inherent in fistula repair surgery) or perioperative procedures and care, are equally or more important in determining fistula closure.

The Lawson and Waaldijk systems fared comparatively worse in terms of predicting fistula closure. Only one component of the Lawson classification system predicted repair outcomes, precluding us from testing the discriminatory value of this system and indicating that fistula location alone may have limited prognostic utility. Few women were reported to have “closing mechanism involvement,” which negatively influenced the performance of the Waaldijk system. Two other studies found the majority of patients to have closing mechanism involvement.^{10,11} This difference may result from varying definitions of “closing mechanism” across surgeons. Nonetheless, Capes and colleagues¹⁵ similarly found that the Waaldijk system performed worse than Goh in terms of prognostic value.

Our analyses indicated potential for simplification of existing systems. For instance, as described by Tafesse, the “Class” subcomponents include different combinations of prior repair and circumferential injury implying a joint effect of these factors on repair outcome which differs from the independent effects of each. When we tested for evidence of multiplicative interaction the cross-product term for prior repair and circumferential fistula was not significant, and the effect estimate for the variable representing “Class 4” fistulas (the joint effect of both factors) was not consistent with the joint effect of both factors being either super-additive or super-multiplicative. Thus, it may be sufficient to account for prior repair or circumferential fistula independently.

Similarly, the Tafesse, Goh and WHO systems have potentially overlapping components. For instance, each includes circumferential fistula and urethral involvement as separate components, though circumferential fistulas involve the urethra. In addition, the WHO component “non-VVF” overlaps with the components measuring urethral involvement and location of the ureters, since the latter are consistent with urethral and ureteric fistulas.

Moreover, several components of the tested systems did not independently predict fistula closure, suggesting they are unnecessary. Ureteric involvement, fistula diameter, mixed RVF/VVF, and cervical fistulas were not statistically significant; prior repair was only marginally significant. These results are similar to what has been previously published.¹⁴ While scarring did not achieve statistical significance after controlling for other components

of the WHO classification system this is likely due to the high degree of correlation between scarring and extensive tissue loss, another component of the system, and the fact that the category includes “mild scarring,” which may not influence repair outcomes.

The empirically-derived prognostic score achieved a discriminatory value similar to the Tafesse, Goh and WHO systems. The score was informed by these systems; however, it included fewer components. Moreover, its components were non-overlapping and used more objective measures, thereby improving likelihood of inter-observer reliability. For instance, we measured “partial urethral involvement” and “circumferential fistula or complete destruction of the urethra” separately, precluding overlap between these components. Similarly, we included measures of scarring presence rather than tissue loss, since it may be easier to measure presence of a factor than its absence. Finally, while for comparison purposes it was necessary to transform the existing systems into scores, none are scoring systems. A prognostic score that is simple and easy to recall, such as the one we tested, could be used in clinical settings to assist surgeons in making decisions about patient triage and planning a repair. Such a score could also be used for research purposes, to facilitate the statistical adjustment for confounding by prognosis of repair, enabling comparison of results across studies.

There are some limitations to this study. Repair outcome was not routinely evaluated via dye test at two study sites; however, at these sites any woman reporting continued incontinence underwent a dye test. As women experiencing incontinence are eager to have this issue rectified, we believe that under-reporting would be minimal. The measures from our data were in some cases approximations of components in classification systems, which may have affected our ability to accurately assess the predictive value of these components. Nonetheless, we approximated these measures to the best of our ability. We also found that model performance declined in the validation cohort compared to the derived cohort. This may be due to the relatively few failed fistula closures and thus unstable estimates. Similarly, use of a split-sample design may have decreased power to detect small effects in the derivation cohort; nonetheless, it also decreased the likelihood of biased measures of classification system performance.²²

In summary, this study empirically evaluated the discriminatory value of existing fistula classification systems for predicting repair prognosis, using data collected from a large and heterogeneous sample of patients across several countries and multiple study sites. While many of the components of the existing genitourinary fistula classification systems we tested predict repair outcomes, none had good prognostic value. Our results also suggest that existing systems could be considerably simplified for prognostic purposes. The prognostic score we empirically derived combines elements of the two most discriminatory systems into a single simpler and more objective measure. These analyses thus represent an important contribution to efforts spearheaded by WHO and other international agencies towards the development and acceptance of a single, standardized, evidence-based fistula classification system. Validation of our findings among other populations of fistula patients and comparison of the inter-rater and intra-rater reliability of these classification systems is warranted.

Acknowledgments

We would like to thank Leslie Davidson, Scott Hammer and Haomiao Jia from Columbia University, Evelyn Landry and Karen Beattie from EngenderHealth, and Erin Mielke, Mary Ellen Stanton and Neal Brandes from USAID, for their thoughtful reviews of the manuscript. These individuals received no compensation for their assistance. Views expressed here do not necessarily reflect those of USAID or the U.S. National Institutes of Health. We are extremely grateful to all of the patients who took part in the study, and to the facility staff who provided care to them and completed case report forms.

Financial support: This work was supported by a Ruth L. Kirschstein National Research Service Award (NRSA) which supported the first author's doctoral research. The overall study was funded by the United States Agency for International Development (USAID), under the terms of associate cooperative agreement GHS-A-00-07-00021-00.

References

1. Sims JM. On the treatment of vesico-vaginal fistula. *American journal of medical science*. 1852; 23:59–82.
2. Creanga AA, Genadry RR. Obstetric fistulas: a clinical review. *Int J Gynaecol Obstet*. 2007; 99(Suppl 1):S40–6. [PubMed: 17868675]
3. Goh J, Stanford EJ, Genadry R. Classification of female genito-urinary tract fistula: a comprehensive review. *Int Urogynecol J Pelvic Floor Dysfunct*. 2009
4. Arrowsmith SD. The classification of obstetric vesico-vaginal fistulas: a call for an evidence-based approach. *Int J Gynaecol Obstet*. 2007; 99(Suppl 1):S25–7. [PubMed: 17869252]
5. Genadry RR, Creanga AA, Roenneburg ML, Wheelless CR. Complex obstetric fistulas. *Int J Gynaecol Obstet*. 2007; 99(Suppl 1):S51–6. [PubMed: 17765241]
6. Wall LL, Arrowsmith SD, Briggs ND, Browning A, Lassey A. The obstetric vesicovaginal fistula in the developing world. *Obstet Gynecol Surv*. 2005; 60:S3–S51. [PubMed: 16034313]
7. Lawson JB. Tropical gynaecology. Birth-canal injuries. *Proc R Soc Med*. 1968; 61:368–70. [PubMed: 5653998]
8. Goh JTW. A new classification for female genital tract fistula. *Aust N Z J Obstet Gynaecol*. 2004; 44:502–4. [PubMed: 15598284]
9. Tafesse B. New classification of female genital fistula. *J Obstet Gynaecol Can*. 2008; 30:394–5. [PubMed: 18505662]
10. Waaldijk K. Surgical classification of obstetric fistulas. *Int J Gynaecol Obstet*. 1995; 49:161–3. [PubMed: 7649321]
11. Raassen TJ, Verdaasdonk EG, Vierhout ME. Prospective results after first-time surgery for obstetric fistulas in East African women. *Int Urogynecol J Pelvic Floor Dysfunct*. 2008; 19:73–9. [PubMed: 17492390]
12. Goh JT, Browning A, Berhan B, Chang A. Predicting the risk of failure of closure of obstetric fistula and residual urinary incontinence using a classification system. *Int Urogynecol J Pelvic Floor Dysfunct*. 2008; 19:1659–62. [PubMed: 18690403]
13. World Health Organization (WHO). *Obstetric Fistula: Guiding principles for clinical management and programme development*. World Health Organization; Geneva: 2006.
14. Frajzyngier V, Ruminjo J, Barone MA. Factors influencing urinary fistula repair outcomes in developing countries: a systematic review. *Am J Obstet Gynecol*. Feb 20.2012 Epub ahead of print.
15. Capes T, Stanford EJ, Romanzi L, Foma Y, Moshier E. Comparison of two classification systems for vesicovaginal fistula. *Int Urogynecol J*. Jan 25.2012 Epub ahead of print.
16. Rothman, KJ.; Greenland, S.; Lash, TL. *Modern epidemiology*. 3rd ed. Wolters Kluwer Health/ Lippincott Williams & Wilkins; Philadelphia: 2008.
17. Carter RE, Zhang X, Woolson RF. C.C. A. Statistical analysis of correlated relative risks. *Journal of Data Science*. 2009; 7:397–407.
18. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol*. 2005; 162:199–200. [PubMed: 15987728]
19. SAS Institute. [Accessed 22 May 2012] Sample 25017: Nonparametric comparison of areas under correlated ROC curves. 2012. 2005at <http://support.sas.com/kb/25/017.html>
20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44:837–45. [PubMed: 3203132]
21. Herschorn S. Female pelvic floor anatomy: the pelvic floor, supporting structures, and pelvic organs. *Rev Urol*. 2004; 6(Suppl 5):S2–S10. [PubMed: 16985905]
22. Gönen, M. *Analyzing receiver operating characteristic curves with SAS*. SAS Institute; Cary, N.C.: 2008.

Figure 1.
ROC curves for each system using the derived cohort

Figure 2.
ROC curves for each system using the validation cohort

Table 1

Creation of variables representing components of the compared classifications.

Classification system	Classification system component	Variable used to operationalize component
Goh	Type 1 Distal edge of the fistula >3.5 cm from external urinary meatus Type 2 Distal edge of the fistula 2.5-3.5 cm from external urinary meatus Type 3 Distal edge of the fistula 1.5-<2.5 cm from external urinary meatus Type 4 Distal edge of the fistula <1.5 cm from external urinary meatus a. Size <1.5 cm in the largest diameter b. Size 1.5-3 cm in the largest diameter c. Size >3 cm in the largest diameter i. None or only mild fibrosis (around fistula and/or vagina) and/or vaginal length >6cm, normal bladder capacity ii. Moderate or severe fibrosis (around fistula and/or vagina) and/or reduced vaginal length and/or bladder capacity iii. Special considerations, e.g. post-radiation, ureteric involvement, circumferential fistula, ⁱ prior	Type 1 Urethral length > 3.5 cm Type 2 Urethral length 2.5-3.5 cm Type 3 Urethral length 1.5-<2.5 cm Type 4 Urethral length <1.5 cm a. Available from dataset b. Available from dataset c. Available from dataset i. None or only mild fibrosis and normal bladder capacity ⁱⁱ ii. Moderate or severe fibrosis and small bladder capacity iii. Ureteric involvement, circumferential fistula, previous repair
Lawson	i. Juxta-urethral ii. Mid-vaginal iii. Juxta-cervical iv. Vault v. Massive combination fistula ⁱⁱⁱ	i. Available from dataset ii. Available from dataset iii. Available from dataset iv. Available from dataset v. Available from dataset
Tafesse	Class 1 Non-circumferential, not previously operated Class 2 Non-circumferential, previously operated Class 3 Circumferential, not previously operated Class 4 Circumferential, previously operated Urethral involvement I. No involvement (length>4cm) II. Urethra involved but not middle 1/3 (length 2.7-3.9 cm) III. Middle 1/3 partly involved (length 1.4-2.6 cm) IV. Middle 1/3 completely involved but some urethral tissue remains (<1.4 cm) V. No urethra	Class 1 Available from dataset Class 2 Available from dataset Class 3 Available from dataset Class 4 Available from dataset Urethral involvement I. Available from dataset II. Available from dataset III. Available from dataset

Classification system	Classification system component			Variable used to operationalize component
	<p>Bladder size</p> <ol style="list-style-type: none"> Longitudinal diameter >7 cm Longitudinal diameter 4-7 cm Longitudinal diameter <4 cm <p>Anterior vaginal tissue loss</p> <ol style="list-style-type: none"> Less than 50% of anterior vagina is involved (>3.5 cm of healthy vagina remains) More than 50% of the anterior vagina wall is involved (<3.5 cm of healthy vagina remains) Obliterated vagina (cannot admit > 1 finger) 			<p>IV. Collapsed categories IV,V</p> <p>Bladder size</p> <ol style="list-style-type: none"> Normal bladder Small or no bladder Small or no bladder <p>Anterior vaginal tissue loss</p> <ol style="list-style-type: none"> Minimal tissue loss Moderate tissue loss Extensive tissue loss or obliterated vagina
Waldijk	<p>Classification</p> <p>Type 1 Not involving the closing mechanism</p> <p>Type 2 Involves closing mechanism</p> <ol style="list-style-type: none"> <i>Without</i> (sub)total urethra involvement <i>With</i> (sub)total urethra involvement <ol style="list-style-type: none"> <i>Without</i> circumferential defect <i>With</i> circumferential defect <p>Type 3 Ureteric and other exceptional fistulas</p> <p>Size</p> <p>Small <2</p> <p>Medium 2-3</p> <p>Large 4-5</p> <p>Extensive 6</p>			<p>Classification</p> <p>Type 1 Not involving the closing mechanism and not involving complete destruction of bladder neck</p> <p>Type 2 Involves closing mechanism or destruction of bladder neck</p> <ol style="list-style-type: none"> Intact or partially damaged urethra Completely destroyed urethra <ol style="list-style-type: none"> Available from dataset Available from dataset <p>Type 3 Mixed vesicovaginal and rectovaginal fistulas, cervical and ureteric fistulas</p> <p>Size</p> <p>Available from dataset</p> <p>Available from dataset</p> <p>Available from dataset</p> <p>Available from dataset</p>
WHO	Defining Criteria	Simple	Complex	
	Number of fistula	■ single	■ multiple	■ Available from dataset
	Site	■ Vesico vaginal (VVF)	■ All non-VVF urinary fistula	■ Non-VVF excludes ureteric and urethral fistulas ^{iv}
			■ recto-vaginal (RVF)	
			■ mixed VVF/RVF	
			■ involvement of cervix	

Classification system	Classification system component			Variable used to operationalize component
	Size (diameter)	■ <4cm	■ >4cm	■ Available from dataset
	Involvement of the urethra / continence mechanism	■ absent	■ present	■ Available from dataset
	Scarring of vaginal tissue	■ absent	■ present	■ Available from dataset
	Presence of circumferential defect	■ absent	■ present	■ Not included in multivariate analysis ^v
	Degree of tissue loss	■ minimal	■ extensive	■ Moderate and minimal tissue loss considered “minimal”
	Ureter/bladder involvement	■ ureters are inside the bladder, not draining into the vagina	■ one or both ureters draining into the vagina ■ one or both ureters at edge of fistula	■ Created composite measure representing ureteric involvement (either ureteric location or ureters draining into vagina or at edge of vagina)
	Number previous repair attempts	■ no previous attempt	■ failed previous repair attempts	■ Available from dataset

ⁱ Complete separation of the urethra from the bladder

ⁱⁱ No measure of vaginal length or bladder capacity available

ⁱⁱⁱ Defined by Lawson as a fistula that is juxta-urethral, midvaginal and juxtacervical

^{iv} Ureteric and urethral fistulas excluded since ureteric fistulas captured under “ureter/bladder involvement” and urethral fistulas captured by “involvement of the urethra/continence mechanism”

^v Excluded from multivariate analysis given overlap with “involvement of the urethra/continence mechanism”

Table 2

Comparison of derived and validation cohort on baseline characteristics and repair outcome

	Total N (%)	Derived Cohort N (%)	Validation Cohort N (%)
Total	1274 (100)	637 (100)	637 (100)
Rural residence	1088 (86.1)	546 (86.4)	542 (85.8)
Mean age	28.2 (11.0)	28.2 (11.1)	28.1 (11.0)
Primary education	267 (21.0)	120 (18.9)	147 (23.1)
Years with fistula	3.3 (5.5)	3.4 (5.6)	3.2 (5.4) *
Previous repair y/n	294 (23.1)	149 (23.4)	145 (22.9)
Type of fistula VVF only RVF and VVF	1229 (97.1) 37 (2.9)	622 (98.3) 11 (1.7)	607 (95.9) ** 26 (4.1)
Current marital status single married/as if married widowed divorced or separated other	23 (1.8) 830 (66.1) 61 (4.9) 341 (27.1) 1 (0.1)	10 (1.6) 403 (64.4) 34 (5.4) 178 (28.4) 1 (0.2)	13 (2.1) 427 (67.8) 27 (4.3) 163 (25.9) 0 (0.0)
Parity	3.4 (2.9)	3.3 (2.9)	3.4 (2.9)
Commodities in residence piped water flush toilet electricity radio TV mobile phone land line phone refrigerator	288 (22.7) 46 (3.6) 003;256 (20.1) 881 (69.2) 199 (15.7) 457 (36.0) 24 (1.9) 49 (3.9)	129 (20.3) 24 (3.8) 119 (18.7) 438 (68.8) 94 (14.8) 221 (34.7) 12 (1.9) 22 (3.5)	159 (25.0) ** 22 (3.5) 137 (21.5) 443 (69.5) 105 (16.5) 236 (37.2) 12 (1.9) 27 (4.2)
Current ability to meet basic needs easily meet needs somewhat meet needs barely satisfy needs	327 (25.8) 660 (52.1) 281 (22.2)	153 (24.2) 336 (53.1) 144 (22.7)	174 (27.4) 324 (51.0) 137 (21.6)
Closed at discharge	1058 (84.7)	534 (85.6)	524 (84.3)
Closed at 3 months	1041 (81.6)	519 (81.5)	522 (82.0)

*
p-value <.05**
p-value <.20

Table 3

Derivation of scores for existing classification systems

Component	Not Closed N (%)	Closed N (%)	RR (95% CI)	ARR (95% CI) ^{vi}	Score ^{vii}
Goh					
Type 1 Distal edge of the fistula >3.5 cm from external urinary meatus (EUM)	20 (18.3)	165 (32.9)	Ref	Ref	-
Type 2 Distal edge of the fistula 2.5-3.5 cm from EUM	47 (46.5)	180 (39.5)	2.58 (1.43- 4.65) **	2.04 (1.60-2.61) **	2
Type 3 Distal edge of the fistula 1.5-<2.5 cm from EUM	27 (26.7)	120 (26.3)	2.43 (1.18-5.03) **	1.68 (1.07-2.66) **	2
Type 4 Distal edge of the fistula < 1.5 cm from EUM	15 (14.9)	37 (8.1)	4.03 (1.90-8.57) **	2.21 (1.33-3.67) **	2
a Size < 1.5 cm	21 (18.4)	107 (21.7)	Ref	Ref	-
b Size 1.5-3 cm	49 (43.0)	273 (55.5)	0.88 (0.57-1.37)	0.74 (0.48-1.12)	-
c Size >3 cm	44 (38.6)	112 (22.8)	1.58 (0.95-2.64) *	0.91 (0.63-1.33)	-
i. None or only mild fibrosis, and/or vaginal length >6cm, normal bladder capacity	Ref	Ref	Ref	Ref	-
ii. Moderate or severe fibrosis, and/or reduced vaginal length and/or bladder capacity	75 (63.6)	216 (41.6)	1.98 (1.22-3.23) **	1.77 (1.19-2.64) **	2
iii. Special considerations, e.g. post-radiation, ureteric involvement, circumferential fistula, previous repair	73 (61.9)	218 (42.0)	1.83 (1.04-3.21) **	1.49 (0.86-2.57)	-
Lawson					
Juxta-urethral	24 (20.5)	105 (20.4)	1.16 (0.85-1.60)	0.95 (0.61-1.46)	-
Mid-vaginal	20 (17.1)	172 (33.2)	0.57 (0.37-0.89) **	0.55 (0.33-0.90) **	-2
Juxta-cervical	20 (17.1)	87 (16.9)	0.94 (0.61-1.46)	0.81 (0.56-1.18)	-
Vault	2 (1.7)	17 (3.3)	0.66 (0.34-1.26) **	0.57 (0.27-1.20) *	-
Massive combination	2 (1.7)	5 (1.0)	1.78 (0.70-4.45)	--	-
Tafesse					
Class 1 Non-circumferential, not previously operated	50 (42.4)	352 (67.8)	Ref	Ref	-
Class 2 Non-circumferential, previously operated	28 (23.7)	98 (18.9)	1.63(0.97-2.73) **	1. 73 (0.93-3.23) *	-
Class 3 Circumferential, not previously operated	29 (24.6)	57 (11.0)	2.58 (1.44-4.63) **	1. 95 (1.05-3.62) **	2
Class 4 Circumferential, previously operated	11 (9.3)	12 (2.3)	3.14 (1.85-5.35) **	2. 28 (1.27-4.11) **	2
I No urethral involvement (urethral length>4cm)	12 (11.9)	116 (25.4)	Ref	Ref	-
II Urethra involved but not middle 1/3 (2.7-3.9 cm)	47 (46.5)	182 (39.9)	2.56 (1.39-4.72) **	1.86(1.27-2.74) **	2
III Middle 1/3 partly involved (1.4-2.6 cm)	34 (33.7)	142 (31.1)	2.60 (1.26- 5.36) **	1.35 (0.67-2.72)	-
IV-V Middle 1/3 completely	8 (7.9)	16 (3.5)	4.46 (1.99-9.98) **	2.17 (1.10-4.29) **	2

Component	Not Closed N (%)	Closed N (%)	RR (95% CI)	ARR (95% CI) ^{vi}	Score ^{vii}
involved or no urethra ^{viii}					
a Longitudinal diameter of bladder > 7 cm	64 (59.3)	368 (75.6)	Ref	Ref	-
b-c Longitudinal diameter of bladder > 7 cm ^{ix}	44 (40.7)	119 (24.4)	1.99 (1.23-3.22) **	1.19 (0.78-1.80)	-
< 50% of anterior vagina involved	34 (28.8)	292 (56.5)	Ref	Ref	-
> 50% of the anterior vagina wall involved	53 (44.9)	190 (36.8)	1.56 (0.99-2.48) **	1.57 (1.21-2.04) **	2
Obliterated vagina	31 (26.3)	36 (7.0)	3.16 (1.99-5.02) **	2.64 (2.17-3.21) **	3
Waaldijk					
Type 1 Not involving closing mechanism	101 (84.9)	490 (94.8)	Ref	Ref	-
Type 2 Involving closing mechanism	-	-	-	-	-
Type 2Aa <i>Without</i> (sub)total urethra involvement without circumferential defect	8 (6.8)	9 (1.7)	2.42 (1.48-4.00) **	2.70 (1.79-4.08) **	3
Type 2Ab <i>Without</i> (sub)total urethra involvement with circumferential defect	6 (5.1)	13 (2.5)	1.89 (0.85-4.19) *	1.67 (0.82-3.37) *	-
Type2Ba <i>With</i> (sub)total urethra involvement without circumferential defect	1 (0.9)	3 (0.6)	1.63 (0.71-3.75)	1.69 (0.70-4.08)	-
Type2Bb <i>With</i> (sub)total urethra involvement with circumferential defect	2 (1.7)	1 (0.2)	3.73 (2.77-5.04) **	3.50 (2.26-5.42) **	4
Type 3 Ureteric and other exceptional fistulas	39 (33.1)	128 (24.7)	1.41 (0.90-2.21) *	1.31 (0.82-2.12) *	-
Small <2	27 (23.7)	143 (29.1)	Ref	Ref	-
Medium 2-3	49 (42.6)	254 (51.7)	0.91 (0.62-1.37)	0.95 (0.65-1.38)	-
Large 4-5	31 (27.0)	75 (15.3)	1.59 (1.06-2.38) **	1.38 (0.97-1.97) *	-
Extensive 6	7 (6.1)	22 (4.5)	1.20 (0.54-2.69)	1.17 (0.61-2.23)	-
WHO					
>1 urinary fistula	16 (13.6)	24 (4.6)	2.12 (1.38-3.26) **	2.13 (1.27-3.56) **	2
Site (mixed vvf rvf or cervical fistula)	8 (6.8)	47 (9.1)	0.74 (0.53-1.04) *	0.83 (0.57-1.21)	-
Size (diameter 4 cm)	38 (33.3)	95 (19.3)	1.66 (1.10-2.50) **	1.13 (0.85-1.51)	-
Involvement of the urethra / continence mechanism	72 (61.0)	192 (37.1)	2.04 (1.52-2.76) **	1.80 (1.28-2.54) **	2
Scarring	94 (79.7)	386 (74.5)	1.30 (0.94-1.80) *	0.99 (0.66-1.48)	-
Circumferential defect ^x	40 (33.9)	69 (13.3)	2.32 (1.64-3.30) **		
Extensive tissue loss	31 (26.3)	35 (6.8)	2.64 (1.83-3.80) **	1.90 (1.38- 2.62) **	2

Component	Not Closed N (%)	Closed N (%)	RR (95% CI)	ARR (95% CI) ^{vi}	Score ^{vii}
Ureter involvement	32 (27.4)	87 (16.9)	1.64 (0.97-2.76) *	1.12 (0.73-1.73)	-
Previous repair	39 (33.1)	110 (21.2)	1.43 (1.01- 2.04) **	1.38 (0.96-1.98) *	-

**
p-value <.05

*
p-value < .20

^{vi}RRs for each component were adjusted for all other components of the classification system

^{vii}Scores were derived by rounding ARR to nearest whole number

^{viii}Categories IV and V were collapsed due to the presence of only 1 woman in the latter category

^{ix}Categories b and c were collapsed because they were equated with “small” or “no bladder” in our dataset

^xThis variable was not included in multivariate analysis since circumferential fistulas are captured under the component “involvement of the urethra.”

Table 4

Additional variables for consideration in a prognostic classification system

Component	Not Closed N (%)	Closed N (%)	RR (95% CI)
Patient characteristics			
Age > 25	65 (55.1)	241 (46.4)	1.10 (0.77-1.56)
Duration of fistula (average years, sd)	5.5 (8.3)	3.0 (4.7)	1.04(1.03-1.06) **
Comorbidities present at baseline			
Genital cutting	35 (29.7)	99 (19.2)	1.31 (0.88-1.95)
Malnutrition	8 (6.8)	31 (6.0)	1.01 (0.46-2.22)
Anemia	9 (7.6)	36 (6.9)	0.88 (0.62-1.24)
UTI	0 (0.0)	2 (0.4)	-
HIV	0 (0.0)	2 (0.4)	-
Malaria	1 (0.8)	3 (0.6)	0.93 (0.33-2.66)
Fistula characteristics and/or categorizations of fistula characteristics not included in above classification systems			
Necrotic tissue present	16 (13.7)	46 (8.9)	1.33 (0.61-2.86)
No or mild scarring	51 (43.2)	356 (68.7)	Ref
Moderate scarring	43 (36.4)	133 (25.7)	1.74 (1.08-2.82) **
Severe scarring	24 (20.3)	29 (5.6)	3.27 (1.91-5.68) **
No urethral involvement	46 (39.0)	326 (62.9)	Ref
Partial urethral involvement	30 (25.4)	119 (23.0)	1.52 (1.12-2.07) **
Complete destruction or transection / circumferential injury	41 (35.0)	72 (14.0)	2.65 (1.87-3.76) **
Non-vvf (ureteric, urethral, rectovaginal, cervical fistula)	78 (66.7)	266 (51.6)	1.71 (1.19-2.44) **
Cervix not visible	27 (22.9)	79 (15.4)	1.43 (0.89 -2.72) *

**
p-value <.05*
p-value <.20

Table 5

Performance of selected classification systems

	Derived cohort		Validation cohort	
Total scores for each scoring system	Proportion failed closures	AUC (95%CI)	Proportion failed closures	AUC (95%CI)
Waalwijk		0.53 (0.51- 0.56)		0.51 (0.49-0.53)
0	108/616 (17.53%)		110/616 (17.83%)	
3	8/17 (47.06%)		3/12 (25.00%)	
4	2/3 (66.67%)		2/5 (40.00%)	
Tafesse		0.66 (0.61- 0.71)		0.60 (0.55-0.65)
0	16/184 (8.70%)		16/188 (8.51%)	
2	38/253 (15.02%)		47/224 (20.98%)	
3	64/200 (32.00%)		52/225 (23.11%)	
Goh		0.62 (0.57-0.67)		0.62 (0.57- 0.68)
0	14/141 (9.93%)		18/141 (12.77%)	
2	44/275 (16.00%)		33/274 (12.04%)	
4	60/221 (27.15%)		64/222 (28.83%)	
WHO		0.69 (0.64-0.74)		0.63 (0.57-0.68)
0	32/337 (9.5%)		44/351 (12.54%)	
2	54/233 (23.18%)		44/215 (20.47%)	
4	32/67 (47.76%)		27/71 (38.03%)	
Empirically-derived		0.70 (0.65-0.75)		0.62 (0.56-0.67)
0	23/277 (8.30%)		32/271 (11.81%)	
1	10/78 (12.82%)		16/177 (20.78%)	
2	29/121 (23.97%)		24/147 (16.33%)	
3	56/161 (34.78%)		43/142 (30.28%)	