TUNEL as a Test for Sperm DNA Damage in the Evaluation of Male Infertility

Rakesh K. Sharma, Edmund Sabanegh, Reda Mahfouz, Sajal Gupta, Aparna Thiyagarajan, and Ashok Agarwal

OBJECTIVES
To standardize the TUNEL assay by establishing inter- and intraobserver variability, interassay variability, cutoff values, sensitivity and specificity of the assay, and studying the distribution of the DNA damage in a population of infertile men referred to a clinical andrology laboratory.

METHODS
Seminal ejaculates from 25 healthy male volunteers (controls) and 194 infertile men (with male factor infertility) referred to an andrology laboratory were examined for DNA damage by TUNEL assay using flow cytometric analysis.

RESULTS
Both the inter- and intraobserver variability and interassay variability was small (<10%). DNA damage in the controls was 11.9 ± 6.8% vs. 29.5 ± 18.7% in patients (P <.001). The cut-off value of 19.25% maximized the observed sensitivity (64.9%) and specificity (100%) of the assay. The distribution of DNA damage in the patients was as follows: 14.9% (29 of 194) with DNA damage between 0% and 10%; 22.7% (44 of 194) between 10% and 20%; 8.8% (17 of 194) between 20% and 30%; and 17.5% (34 of 194) between 30% and 40%. Finally, 27.3% (53 of 194) had TUNEL values >40%.

CONCLUSIONS
We report a detailed standardization of the TUNEL assay for clinical use, as well as reference ranges for DNA damage in normal healthy donors and infertile men. A cut-off of 19.25% with observed 100% specificity established in our program can differentiate infertile men with DNA damage from healthy men. This test can be offered to infertile patients who are idiopathic, have severe oxidative stress-related abnormal semen quality, and contribute to the infertility problem of the couple who are considering assisted reproductive techniques.

E
evaluation of sperm DNA damage can better predict the outcome of spontaneous pregnancy and assisted reproductive techniques (ART) than traditional sperm parameters. Studies have highlighted the usefulness of measuring DNA damage in natural and ART settings.

The 2 most commonly used tests to measure sperm DNA damage are the terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) assay and the sperm chromatin structure assay (SCSA). DNA testing by SCSA has been extensively standardized and it has been performed by a single primary reference laboratory (SCSA Diagnostics, Brookings, SD) or its approved centers. Other laboratories have had difficulties in standardizing this assay. TUNEL test has not been standardized to the same extent as SCSA and few laboratories have attempted to use this test in a clinical setting, limiting its use in patients in past. The TUNEL assay measure both single- and double-strand DNA fragmentation, measures a definitive end point (presence of free 3’ hydroxyl groups), and can provide more meaningful information on the implantation potential of an embryo. It is technically less demanding, unlike SCSA assay, where semen samples are batched and shipped to the primary SCSA testing laboratory and analyzed for DNA damage using a dedicated flow cytometer. However, the TUNEL assay has never been optimized for use with human spermatozoa and lower normal threshold values have not been clearly established. Our goal was to standardize the TUNEL assay and establish a platform for andrology laboratories with access to flow cytometry facilities, where the clinician or the urologist can offer this test to infertile men with poor sperm quality because of oxidative stress and/or patients presenting with idiopathic infertility. Specifically, we measured the inter- and intraobserver variability, inter- and intraassay variability, and we established a cut-off value, sensitivity, and specificity of sperm DNA fragmentation levels in healthy men and in patients with male factor infertility.

MATERIAL AND METHODS

Subjects Selection
This study was approved by our Institutional Review Board. We enrolled 194 infertile men in our study from March 2008 to December 2009. All infertile patients were seeking medical...
advice for male factor infertility (idiopathic infertility, varicoceles, infection, and other known etiologies). Twenty-five healthy male volunteers (controls) of proven (n = 7; initiated a pregnancy within the past 1 year) and unproven fertility (18; normal semen analysis results but had not established a pregnancy) were selected on the basis of normal semen analysis according to guidelines of the World Health Organization.8

Semen Collection and Preparation
Semen specimens were collected by masturbation and a 5-μL aliquot was loaded on a 20-μL Cell-Vu chamber (Millennium Sciences, Inc., New York, NY). Manual semen analysis and morphology was done using both the World Health Organization’s and Kruger’s strict criteria.8 An aliquot of well-liquefied seminal ejaculate was used to assess DNA damage using the TUNEL assay. Briefly, 1-2 million spermatozoa were washed in phosphate-buffered saline (PBS) and resuspended in 3.7% paraformaldehyde. They were placed on ice for 30-60 minutes at 4 °C. Thereafter, the spermatozoa were again washed to remove the paraformaldehyde and then re-suspended in 70% ice-cold ethanol and stored at −20 °C until the run time.

Sperm DNA Fragmentation by Terminal Deoxynucleotidyl Transferase-Mediated Fluorescein-TUNEL Assay
Sperm DNA fragmentation was evaluated using a terminal deoxynucleotidyl transfrase–mediated fluorescein-TUNEL assay with an Apo-Direct kit (Pharmingen, San Diego, CA) as described earlier.9,10 Positive and negative kit controls provided by the manufacturer and internal controls (specimens from donors and patients with known DNA damage) were included for each run. After a second wash in PBS to remove ethanol, the sperm pellets were resuspended in 50 μL of freshly prepared staining solution for 60 minutes at 37 °C. The staining solution contains terminal deoxynucleotidyl transferase (TdT) enzyme, TdT reaction buffer, fluorescein isothiocyanate–tagged deoxyuridine triphosphate nucleotides (FITC-dUTP), and distilled water. All specimens were further washed in rinse buffer, resuspended in 0.5 mL of propidium iodide/RNase solution, and incubated for 30 minutes followed by flow cytometric analysis.

Flow Cytometry Analysis
All fluorescence signals of labeled spermatozoa were analyzed by the flow cytometer FacScan (Becton Dickinson, San Jose, CA). About 10,000 spermatozoa were examined for each assay at a flow rate of <100 cells/s. The excitation wavelength was 488 nm supplied by an argon laser at 15 mW. Green fluorescence (480-530 nm) was measured in the FL-1 channel and red fluorescence (580-630 nm) in the FL-2 channel. The percentage of positive cells (TUNEL-positive) was calculated on a 1023-channel scale using the flow cytometry software FlowJo Mac version 8.2.4 (FlowJo, LLC, Ashland, OR).9,10

Interobserver and Intraobserver Variation
Two observers (RKS and RM) evaluated interobserver and intraobserver variability on 47 TUNEL measurements. The interobserver variability was obtained by analyzing the differences in the results produced by the 2 observers. The intraobserver variation was obtained by analyzing the differences between measurements within each observer.

Interassay and Intra-assay Variation
Interassay and intra-assay variation were measured using the positive and negative controls provided in the assay kit and samples from the patient population. A total of 18 readings for both negative and the positive controls and 80 readings for the patients were examined.

Statistical Analysis
All analyses were performed using R version 2.9.0.11 Summaries of the categorical variables were reported as percent frequency, and quantitative variables are represented as mean ± standard deviation and median (25th and 75th percentiles). The difference in distributions of TUNEL-positive levels between infertile patients and donors was assessed using the Wilcoxon rank-sum test. A P value <.05 was considered statistically significant. To establish the clinical utility of the TUNEL assay, a receiver operating characteristic (ROC) curve was constructed to obtain the cutoff point for TUNEL-positive sperm. Sensitivity and specificity, positive predictive value (PPV) and negative predictive value (NPV), and accuracy of the test were calculated.

RESULTS

Intraobserver Variation
Because an assay itself has some inherent intra-assay variability, we should be aware that intraobserver variability described in this process includes an element of intra-assay variability. A very practical way to describe the intraobserver variability is to determine the degree to which an individual measurement differs from the final assigned TUNEL measurement. These differences can be constructed as absolute differences or percent differences relative to the assigned values. Both the absolute differences and the percentage differences relative to the assigned values were calculated. When absolute values were calculated, 80.0% of individual TUNEL measurement differed from the final designated values by no more than 3.2%. Similarly 85.7% of individual measurements were within 5.0% of the final designated values (Figure 1A). When the percentage differences were examined, 57.1% of individual measurements in these data had a percentage difference less than 10% of the assigned value (Figure 1B).

Interobserver Variation
The interobserver variability was assessed by the likelihood that an observer’s designated value (for either observer) for a given donor is within a specified value (either by absolute difference or percent difference) from the mean designated value of the 2 observers. The absolute difference between an observer’s designated value and the mean among the 2 observers was within 4.0% in 83.3% of donors (Figure 1C). Similarly, in 83.3% of the donors, the percentage difference between an individual observer’s designated TUNEL value and the 2 observers’ average value was within 15% (Figure 1D).

Intra-assay and Interassay Variability
For both the negative and positive controls, an average of readings 1 and 2 was used, and the likelihood that an
individual measurement would be within specified thresholds of the mean was considered. For the negative controls, both the absolute and percent difference between an individual measurement and the mean of the duplicate measurements was within <1.0% in 88.9% of cases (Figure 2A, B). For the positive controls, the absolute difference from the mean of the duplicate measurements was within 2.0% in 77.8% of cases, and the percentage difference was within 5.0% in 77.8% of cases (Figure 2C, D).

For the patient samples, the absolute difference between an individual measurement and the mean of the pair of measurements was within 2.0% in 86.3% of cases. The percentage difference was within 10% in 90.0% of cases (Figure 2E, F).

Measurement of DNA Damage in Donors and Patients

Of the 25 donors tested, 7 were of proven fertility and 18 were of unproven fertility. When donors and patients were combined, the DNA damage as mean ± standard deviation was 26.3% ± 18.4%; the median (25th and 75th percentiles) was 21.8% (12.3%, 37.8%). DNA damage was significantly small in the donors 11.9% ± 6.8%, 12.6% (6.3%, 16.9%) compared with the patients 29.5% ± 18.7%, 27.5% (14.4%, 41.0) (P < .01). The cut-off value of 19.25% maximized the sum of the sensitivity and specificity. The ROC curve showed an observed sensitivity of 64.9% and observed specificity of 100%, with 0.82 area under the curve (Figure 3). At this cut-off, the positive predictive value was 97.7% and the negative predictive value was 37%, with an accuracy of 70%.

When the DNA damage was categorized into 0%-10%, 10%-20%, 20%-30%, 30%-40%, and >40%, 28.7% of all the subjects had DNA damage within 10%-20%. In the donors, 37.2% had DNA damage in the range of 0%-10% and 55.8% had damage from 10%-20%. Among the patients, only 14.9% (29 of 194) had DNA damage <10% and 22.7% (44 of 194) between 10% and 20%, 8.8% (17 of 194) between 20% and 30%, and DNA damage between 30% and 40% was seen in 17.5% (34 of 194) of the patients, whereas >40% DNA damage was seen in 27.3% (53 of 194) of the patients. DNA damage and relative frequency of DNA damage in donors and patients is shown in Figure 4 (A, B). One-hundred percent of donors and 64.9% of patients had DNA damage below the cut-off value of 19.25.
In most of the samples, only $\sim 5\%$ of sperm have TUNEL-detectable DNA damage. In in vitro fertilization (IVF) patients, lower pregnancy rates (19\%) were reported when the sperm DNA damage rates as determined by TUNEL were above the cut-off value of 36.5%. In intracytoplasmic sperm injection (ICSI) couples, a threshold DNA damage rate of 24.3\% was associ-
ated with a lower chance of pregnancy. Furthermore, no pregnancies were reported above 20% DNA fragmentation. In natural conception, a higher percentage of sperm DNA damage was found in patients with various andrological diseases compared with normozoospermic subjects, as assessed by the TUNEL assay using microscopic evaluation. Similarly, in intrauterine insemination, none of the semen samples with >12% sperm DNA fragmentation resulted in a pregnancy. Information is lacking on the clinical value of the TUNEL assay in predicting pregnancy outcome in vivo.

The goal of the current study was to standardize the TUNEL assay by establishing inter- and intra-observer variability, interassay variability, cutoff value, sensitivity, specificity of the assay, and study the distribution of the DNA damage in a population of infertile men referred to the clinical andrology laboratory. These patients typically present with either idiopathic infertility or have oxidative stress (poor semen quality, leukocytospermia/infection) and may benefit from antibiotics and antioxidants. Furthermore, some of these couples may be considering ART or may have had previous ART failure(s). In analyzing the variations both in the observer and the assay, we chose the actual differences as well as percent differences rather than adopting the variance components approach to these analyses. The reasons for this approach are 2-fold. The first is that these are much more interpretable in terms of actual measurements. The second reason is that the reliability estimates based on variance components are ultimately highly dependent on the total variability in TUNEL measurements on which the analysis is based, and that could change dramatically based on the population being sampled. For example, the total variation among a set of donors would tend to be much less than the total variability among a set of patients, so even if separate observers provided a similar degree of interobserver variability in the 2 populations with respect to their absolute difference in assigned TUNEL values, the results would look more impressive in the more variable patient sample, because the variance components are compared with total variability.

Similarly, in establishing the cut-off or normal threshold values, as well as the sensitivity, specificity, NPV, and PPV, it is important to consider whether a test is to be used as a screening/diagnostic test or as a predictor of an established endpoint. High sensitivity is important for a screening or diagnostic test so that it can be offered to a larger population. However, specificity becomes critical if a test is to be offered as a predictive marker of a defined endpoint. Our results show a very high specificity and a slightly lower sensitivity of the TUNEL assay. Estimated sensitivity values tended to be lower because a fair number of patients had TUNEL measurements in the same range as the donors. However, this type of cutoff is ideal, because anything above the cut-off would seem to strongly suggest infertility. Positive and negative predictive values are dependent on the prevalence of infertility in the tested population and will be different in populations where the percentage of fertile subjects may be higher.

Sergerie et al. reported threshold values of TUNEL in 47 men with proven fertility and 66 infertile men. The infertile men had a higher mean level of DNA damage than the proven fertile men (40.9% ± 14.3% vs 13.1 ± 7.3%) (P < .001). The area under the curve was 0.93 for a 20% cutoff; the specificity was 89.4% and the sensitivity was 96.9%. The PPV and NPV were 92.8% and 95.5%, respectively. We had a larger number of infertile patients (194) vs. 66 in their study.

In our study, the control population consisted of 7 men with proven and 18 with unproven fertility. The cut-off values were close to those reported by Sergerie et al. (19.2% vs 20%). If the goal is to use this test for accurately predicting infertility because of DNA damage, then specificity is more important. In our study, the sensitivity was therefore lower than that reported in the Sergerie study (64.9% vs 96.9%), but the observed specificity was 100%. Similarly, the PPV was high in our study, indicating that sperm DNA fragmentation measured by the TUNEL assay is a good predictive parameter to identify an infertile population. These values are similar to those reported earlier (20% by Benchab et al., 24.3% by Henkel et al., and 24% by Sergerie et al.). These values are much lower than the threshold established for SCSA (≥ 30%). Both the sensitivity and specificity are related with the intrinsic performance of the TUNEL. However, the PPV and NPV are strongly associated with the prevalence of the disease.

To explain this, if the prevalence of infertility in an ART setting is assumed to be 50%, under similar TUNEL assay conditions and similar sensitivity and specificity conditions, the PPV would remain above 95%, indicating that the subject would be falsely marked as infertile in less than 5% of the cases (false positive). It is more acceptable to diagnose subfertile males falsely as fertile instead of diagnosing fertile males as infertile. This
is important because this will reduce over-treatment of potentially fertile couples and the number of them who are referred to ICSI instead of IVF. The goal of our study was to examine whether the TUNEL assay can identify men who will contribute to the infertility problem of the couple rather than a test that classifies them as possibly fertile or subfertile. The predictive value of assessing sperm DNA damage can vary depending on the sperm DNA test and the cut-off level that is used. This is reflected in various studies that have used various cutoff levels in predicting pregnancy loss using the SCSA or TUNEL assay and the odds ratio in predicting pregnancy loss.6

The clinical value of sperm DNA testing has been excellently illustrated in pro and con statements based on a systematic review of literature, sperm DNA characteristics as well as the prevalence of the disease in natural and IUI as well as in IVF and ICSI candidates.5 This study also illustrates the value of PPV vs. NPV over the odds ratio, as well as the prevalence of the disease (pregnancy, pregnancy loss).

A modified protocol for the TUNEL assay has been recently published by Mitchell et al.22 using dithiothreitol to increase the accessibility of terminal transferase to access the sites of DNA cleavage. In addition, an inclusion of the vital stain makes the test more robust and simple and may further help in our understanding of the causes of DNA damage in human spermatozoa. Unfortunately, none of the currently available assays for measuring DNA damage can selectively differentiate clinically important DNA fragmentation from clinically insignificant fragmentation. The assays also cannot differentiate normal (physiological) DNA nicks from pathologic nicks. In addition, the assays do not give any information regarding the particular genes that may be affected by DNA fragmentation. These assays can only determine the amount of DNA fragmentation that occurs, with the assumption that higher levels of DNA fragmentation are pathologic. Additional studies are needed to establish the lower thresholds of the TUNEL assay and evaluate the predictive value of sperm DNA damage.

CONCLUSIONS

We report for the first time a detailed stepwise standardization of the TUNEL assay for clinical use and establish references ranges for DNA damage in normal healthy donors and infertile men. A cut-off of 19.25% with an observed 100% specificity established in our program can differentiate infertile men with DNA damage who are contributing to the infertility problem of the couple. Other centers can use these reference values to develop their own reference cutoff values. The TUNEL test can be offered to a select group of patients presenting with idiopathic infertility or in cases where oxidative stress may be an underlying issue. Use of this test can be effective in establishing the DNA integrity of the sperm in cases of selected male infertility by any fertility testing facility with access to flow cytometry before considering other more expensive ART procedures.

Acknowledgments. The authors wish to thank Debbie Garlak, MT, and Lisa Sullivan, MT, BS from the Andrology Laboratory for their assistance in scheduling the donors; Jeff Hammel, M.S. for his help with the statistical analysis. This study was supported by the Center for Reproductive Medicine, Glickman Urological and Kidney Institute, Cleveland Clinic.

References


