Another look at human sperm morphology

J. Auger1,2,*, P. Jouannet3, and F. Eustache2,4

1Service d’Histologie-Embryologie, Biologie de la Reproduction / CECOS, Hôpitaux Universitaires Paris Centre, Site Port-Royal, 53, Avenue de l’Observatoire, 75014 Paris, France 2INSERM U1016, Equipe “Génomique, Épigénétique et Physiologie de la Reproduction”, Institut Cochin, Université Paris Descartes, Paris, France 3Université Paris Descartes, Paris, France 4Service d’Histologie-Embryologie, Cytogénétique, Biologie de la Reproduction / CECOS, Hôpitaux Universitaires Paris Seine-Saint-Denis, Site Jean Verdier, 93143 Bondy Cedex, France

*Correspondence address. Tel: +33-1-58-41-37-33; Fax: +33-1-58-41-15-65; E-mail: jacques.auger@aphp.fr

Submitted on June 17, 2015; resubmitted on August 29, 2015; accepted on September 15, 2015

STUDY QUESTION: Can a standardized assessment of abnormal human sperm morphology provide additional useful information by identifying men with more severe disturbances in different types of abnormalities?

SUMMARY ANSWER: Definition-based categorization of sperm head, midpiece and tail defects has shown how differently these abnormalities are distributed in fertile men and other groups of men, thus providing high and low thresholds, a starting point for diagnosis or research purposes.

WHAT IS KNOWN ALREADY: Several recent studies have reported indisputable genetic origins for various sperm defects. A few studies have also identified associations between environmental factors and low percentages of morphologically normal spermatozoa. Nevertheless, with the exception of rare situations in which the vast majority of spermatozoa have specific, easily characterized defects, such as ‘globozoospermia’, little attention has been paid to the description and precise quantification of human sperm abnormalities. The lack of standardization in the phenotyping of sperm morphological defects by conventional microscopy is a limiting factor for diagnosis and for intra- or inter-observer or centre consistency in studies investigating the causal factors and possible functional consequences of the abnormalities detected. There are currently no baseline data for abnormalities of sperm morphology based on a standardized classification, in the general population, among fertile or other groups of men.

STUDY DESIGN, SIZE, DURATION: This study is based on detailed sperm abnormality datasets acquired by a standardized classification method, from several groups of men, over the same 5-year period.

PARTICIPANTS/MATERIALS, SETTING, METHODS: We studied cross-sectional data from fertile men (n = 926), male partners from infertile couples (n = 1747) and testicular cancer patients (n = 239). We used a standardized classification to analyse Shorr-stained slides, taking into account all the abnormalities encountered.

MAIN RESULTS AND THE ROLE OF CHANCE: Most sperm defects were significantly more frequent in infertile than in fertile men, with 20–30% of infertile men having frequencies of abnormalities above the 95th percentile in fertile men for 9 out of the 15 categories of abnormalities. Interestingly, several head abnormalities were significantly more frequent in patients with testicular cancer than in infertile men, highlighting the particular impact of this condition on sperm morphogenesis. We used the 95th percentile in fertile men as the lower threshold and the 99th percentile in infertile men as an upper extreme threshold, for the classification of morphological abnormality frequencies into three levels: low, intermediate and high. The assessment of several semen samples, with or without a genetic background, for abnormal sperm morphology, based on the percentage of normal spermatozoa, a teratozoospermia index, and the detailed profile of abnormalities categorized according to the three levels proposed, has highlighted the value of detailed phenotyping for diagnosis and research purposes.

LIMITATIONS, REASONS FOR CAUTION: The thresholds proposed for the various categories of sperm abnormality should be considered relative rather than absolute, owing to the known sampling error related to the limited number of spermatozoa assessed per sample, or when studying the general population or populations from regions other than Western Europe. The standardized assessment of abnormal sperm morphology requires time and experience. We therefore suggest that this assessment is carried out during a first andrological check-up or for epidemiological or research studies, rather than in the routine management of infertile couples for assisted reproductive technologies.

WIDER IMPLICATIONS OF THE FINDINGS: The study design used for the fertile group of men was similar to that previously used for the WHO reference values, providing a rationale for considering the 95th percentile in fertile men as the level below which abnormalities may be considered to occur at a frequency representing random background variations of a normal spermiogenesis process. The crude frequencies obtained, and the
Abnormal sperm morphology

Introduction

Sperm morphology assessment, on a stained semen smear, reveals a plethora of morphological variations or abnormalities of the sperm head, mid-piece and tail components. These abnormalities primarily reflect the complexity of terminal sperm differentiation following several biochemical and morphological changes (acrosome formation, flagellar development, chromatin condensation and major reorganization of the nucleus and cytoplasm) and the influence of multiple genetic or micro- or macro-environmental factors modulating or disrupting this crucial stage in morphogenesis.

Normal and abnormal sperm cells exhibit a number of features on stained smears studied by conventional microscopy. Various sperm morphology classifications have been proposed from the early 1950s (notably with the remarkable contribution of MacLeod and Gold, 1951) until the 1970s (reviewed in Mortimer and Menkveld, 2001) for categorizing spermatozoa according to these various aspects. Spermatozoa retrieved from the uterine cavity or in the vicinity of oocytes recovered from the female reproductive tract, and thus clearly with the potential to migrate in the female genital tract and to bind to oocytes, were found to be more homogeneous in appearance than those from a native semen sample. This observation helped to define the appearance of potentially fertilizing (morphologically normal) spermatozoa (Menkveld et al., 1990; Liu and Baker, 1992). With the development of assisted reproductive technologies (ART) from the 1980s onwards, initially through IVF and then with the development of ICSI in the 1990s, sperm morphology analysis became important for prognostic reasons, but was mostly restricted to assessments of the proportion of normal spermatozoa (Ombelet et al., 1997). The sperm morphology assessment carried out primarily in the context of ART is generally considered to constitute the cornerstone of such analyses in humans. These methods have been fully endorsed in the WHO laboratory manual for the examination and processing of human semen, as the reference method for assessing human sperm morphology (WHO5: World Health Organization, 2010).

Several studies have identified genetic causes of various morphological abnormalities of spermatozoa, such as globozoospermia and macrocephalic sperm, when these abnormalities concern the vast majority of the sperm cells present (Guichaoua et al., 2009), and associations between genetic, lifestyle or environmental factors and abnormal sperm morphology have also been reported (Wyrobek et al., 1983; Buck Louis et al., 2014). However, the precise origins of many of the human sperm abnormalities observed are far from well understood, and the precise impact of these abnormalities on sperm fertility potential also remains unclear.

As a consequence, it has recently been suggested that precise analyses of the number of sperm abnormalities by conventional microscopy would be a useful approach for diagnosis or research purposes (Menkveld, 2013; Mitchell et al., 2015). The WHOS report (World Health Organization, 2010) has come to a similar conclusion, with its statement that ‘categorizing all abnormal forms of spermatozoa may be of diagnostic or research benefit’. WHOS contains a rich iconography on morphological abnormalities of spermatozoa, with legends providing a simple conclusion: ‘normal’ or ‘abnormal’ spermatozoon, with comments in some cases, such as ‘pyriform’ and ‘amorphous’ head. This document does not propose clear definitions of the various abnormalities of human sperm cells, and the corresponding ‘diagnostic or research benefits’ are neither demonstrated nor discussed. This limits the morphological evaluations of spermatozoa and comparisons of the results obtained by different observers or in different studies (Wang et al., 2014).

We have previously reported and used a method for categorizing normal sperm morphology and the various sperm defects encountered (Auger, 2010), in line with contemporary knowledge of sperm assembly (Chemes and Rawe, 2003), the correspondence between the defects observed on conventional microscopy and transmission electron microscopy (Bisson and David, 1975), and the functional consequences of most of the defects categorized (Chemes and Alvarez Sedo, 2012).

One of the principal advantages of such a standardized categorization is the provision of basic data about the distributions of various sperm defects.

This study, based on our standardized method, was designed to establish reference ranges for morphological abnormalities of human spermatozoa, through the analysis of three large groups of men: fertile men, infertile men and men with testicular germ cell tumours (TGCTs). Thresholds for future clinical, epidemiological or basic studies are proposed and discussed.

Materials and Methods

Subjects studied

Sperm morphology was assessed in three groups of men: a homogeneous group of fertile men and two different groups of men of unknown fertility status (male partners from infertile couples, referred to hereafter as ‘infertile men’ for convenience, and men with testicular cancer).

The fertile men were the male partners of pregnant women participating in a study investigating possible geographical differences in semen quality (Jørgensen et al., 2001). Each of the men provided a single semen sample. They were recruited from four different European cities in 1997–1998. These men had a mean age of 31 years (interquartile range (IQR): 24–38 years old) and their partners had become pregnant without the need for fertility treatment (e.g. stimulation of ovulation, ART). In the protocol of this international study, sperm morphology was analysed in a centralized manner, in our laboratory, to minimize the well-known wide variability in human morphology assessment by different observers and centres (Ombelet et al., 1998; Eustache and Auger, 2003). A first analysis focusing on geographical variation in sperm morphology and the influence of environmental and lifestyle factors has already been published (Auger et al., 2001), but the distributions of the various sperm abnormalities were not reported. In this study, only sperm morphology was assessed for couples for whom a pregnancy was achieved within 1 year (n = 926), in accordance with the three levels of abnormality frequency proposed for each standardized category of sperm defect, provide baseline data useful for diagnosis and a starting point for future studies aiming to identify associations with genetic or environmental factors.

STUDY FUNDING/COMPETING INTEREST(S): Part of this study was supported by contract BMH4-CT96-0314 from the European Union. The authors have no competing interests to declare.

Key words: fertile men / human semen / infertility diagnosis / multiple anomalies index / reference values / sperm defects / standardization
methodology used for the definition of the new WHO reference semen values (Cooper et al., 2010).

The second population study consisted of male partners from infertile couples referred to the reproductive biology laboratory of Cochin Hospital Paris for routine semen analysis during the same time period. This group of 1747 men, for whom it was not possible to collect information about the duration of infertility or other clinical information in a systematic manner, was older than the fertile men (mean: 36.5; IQR: 32–40). The dataset consisted of the results of all consecutive routine semen analyses, with only the first semen analysis used in cases of repeated analyses for the same patient during the period.

The third group of men consisted of patients referred to the reproductive biology laboratory of Cochin Hospital Paris by their urologist or oncologist, during the same time period, for semen cryopreservation before anti-neoplastic treatment for a TGCT. Semen was collected before orchidectomy (Petersen et al., 1999). These patients were younger than the men in the other two groups (mean: 29.2; IQR: 25–33).

Pre-analytical steps in sperm morphology assessment

All semen smears were prepared, fixed and stained according to the same standardized procedure. Briefly, smears were prepared from a 10 μl well-mixed drop of semen, which was allowed to dry in air, and then fixed by incubation for 1 h with a mixture of absolute ethanol (2/3) and acetic acid (1/3), before Shorr staining (World Health Organization, 1992) by an automated method (Sakura DR5601, Bayer Diagnostics, Puteaux, France) ensuring the homogeneous staining of different preparations.

Methodology used for human sperm morphology classification

The method used here for the assessment of human sperm morphology was essentially the original method of classification described by David et al. (1975) and then modified in the late 1990s to provide clear rules for the classification of normal and abnormal spermatozoa and for the categorization of the various sperm defects observed in stained smears through conventional microscopy (Auger et al., 2001). The various morphological categories are shown in Fig. 1, with a summary of the definitions in the figure legend. According to this method, morphologically normal spermatozoa are categorized on the basis of the same criteria proposed by WHO from the third edition of the manual to the recent WHOS (World Health Organization, 1992, 2010, respectively). However, unlike the WHOS guidelines, our classification considered all borderline aspects to be normal, as in previous classification systems (Freund, 1966; Eliasson, 1971) and as argued for in previous studies (Auger, 2010). Borderline aspects introduce marked uncertainty into classification systems and are a major component of the overall intra- and inter-observer variability in assessments without categorization rules. In the method used here, a reticle eyepiece with a micrometre was used whenever the observer had doubts about sperm size, in terms, for example, of the lengths of the major axis or the minor axis of the head, and decision-support criteria were defined for distinguishing borderline aspects of size and shape (see Supplementary data, Fig. S1). This method was initially developed for diagnostic purposes. It therefore categorizes the head, midpiece and principal piece defects most frequently observed in human sperm and known to correspond to particular aetiologies or ultrastructural alterations (Chemes and Rawe, 2003). The use of a multi-entry grid, accounting for all abnormalities and their combinations, without underestimating one abnormality with respect to the others (e.g. a sperm cell can have residual cytoplasm in excess as well as a bent tail), makes it possible to calculate an overall indicator of the level of teratozoospermia, the Multiple Anomalies Index (MAI), equal to the mean number of abnormalities per normal spermatozoon (Jouannet et al., 1988; World Health Organization, 2010). Our classification considers true excess residual cytoplasm (ERC), but not physiological cytoplasmic droplets (Cooper, 2005), as abnormal.

Analytical procedure

All stained smears were assigned to a pool of four experienced technicians who assessed them unaware of sample origins, by using the combination of a ×100 oil-immersion bright-field objective and ×10 eyepiece (×1000 final magnification). Each of these technicians routinely assesses two to five stained smears daily, and their accuracy and reproducibility in assessments of human sperm morphology, based on the classification described here, are regularly proved to be consistent (an example for IQC is provided in Supplementary data, Table S1).

Statistical methods

For the three groups of men studied, only data from stained smears for which at least 100 spermatozoa could be classified were analysed. We also compared each abnormality category within each group of men, between men with and without normal sperm production and progressive motility, as defined in WHOS (normozoospermic men: at least 39 million spermatozoa per ejaculate and at least 32% progressively motile sperm cells; World Health Organization, 2010). Age is not associated with sperm morphology, at least for the 25–41 years age group (Schwartz et al., 1984). We also found no association between age and the extent of normal or abnormal morphology in the three groups of men studied (data not shown). We therefore did not adjust comparisons for age. All statistical comparisons between fertile men, infertile men and TGCT patients were performed with BMDP statistical software (Statistical Solutions, Cork, Ireland). The similarity of the mean values for the percentages of normal spermatozoa, all morphological defects and MAI values between the three groups was assessed by one-way analysis of variance (BMDP 7D subroutine), taking unequal variances into account (Brown–Forsythe test) when necessary. When the null hypothesis was rejected, post hoc Tukey tests were used for pair-wise comparisons between groups of men. Crude data were used to determine the distribution profiles for each morphological abnormality category, MAI and morphologically normal spermatozoa in the three groups of men studied. The curves best fitting each distribution profile were then constructed with the Fit-curve option of Sigmaplot software (Statistical Solutions).

By analogy with the WHO reference thresholds for semen characteristics obtained for a similar population of fertile men (Cooper et al., 2010), we considered the 95th percentile for the various categories of abnormalities and MAI, and the 5th percentile for the percentage of normal spermatozoa in the fertile group as lower reference thresholds. Dichotomous categorization has been shown to be of little overall clinical relevance for semen variables (Björndahl, 2011). We therefore also used a second, experience-based, higher threshold, set at the 99th percentile for the population of infertile men, as an extreme level for the sperm abnormalities considered, potentially with a specific origin and possible pathological implications. We then defined three levels of sperm abnormalities on the basis of these two thresholds: a low level of abnormality (LLA), below the 95th percentile in fertile men; a high level of abnormality (HLA), above the 99th percentile in infertile men and an intermediate level of abnormality (ILA), between the 95th percentile in fertile men and the 99th percentile in infertile men. This ILA is probably non-random, particularly if it tends to increase towards extreme values. A preliminary validation of the utility of the three levels of abnormality proposed was made on a sample of six men, one of whom was fertile, the other five having specific teratozoospermia profiles with a known genetic background.

Ethical approval

The study was approved by the Cochin University Hospital for the three groups of French men studied and by all collaborating institutions for the fertile men from other European countries. Informed consent was obtained from all participants.
Figure 1  Standardized classification of normal and abnormal human sperm morphology. Morphologically normal spermatozoa are defined as in the 5th WHO manual (World Health Organization, 2010) except that all borderline aspects are considered normal (decision rules for borderline aspects are presented in Supplementary data, Figure S1); tapered head: increased major axis length and normal minor axis length; thin head: normal major axis length and decreased minor axis length; microcephalic: decreased major and minor axes lengths; macrocephalic: increased major and minor axes lengths; multiple heads: more than one head per sperm cell; abnormal post-acrosomal region: any type of outline and/or size and/or texture abnormality of the posterior part of the head; abnormal acrosomal region: any type of outline and/or size and/or texture abnormality of the anterior part of the head; abnormal (or excess) residual cytoplasm (or abnormal cytoplasmic remnant): residual cytoplasmic area > 30% of the head area, in any part of the sperm cell; thin midpiece: midpiece diameter, smaller than the principal piece diameter for > 40% of the whole midpiece length; bent/ misaligned tail: may include neck anomalies, misalignment of the head major axis and midpiece axis or an acute angle (≤120°) between the head major axis and midpiece axis or an acute angle (≤ 120°) between the midpiece and the principal piece axes; no tail: isolated head, no tail observed; short tail: a tail length (midpiece + principal piece) no more than five times the major axis length of the head; irregularly shaped tail: irregular or changing calibre along the tail; coiled tail: completely or partially coiled tail, with the coil close to or around the head (flat coiling or coiling at the extremity of the tail reminiscent of hypo-osmotic coiling aspects should not be included in this category); multiple tails: more than one tail per sperm cell. Many of the spermatozoa presented more than one type of abnormality. The classification method encompasses all the abnormalities of each cell and the calculation of multiple anomalies index (MAI) (World Health Organization, 2010).
Results

Fertile men

Distributions of morphologically normal spermatozoa and of the various categories of sperm abnormality percentage and of MAI in the 926 fertile men are presented in Table I. The 5th percentile for the percentage of normal spermatozoa including all borderline aspects was 23% (95% CI, 20–26). Head abnormalities were more frequent than midpiece and principal piece defects. Texture or outline abnormalities of the acrosomal or post-acrosomal regions were markedly more frequent among head abnormalities than head size abnormalities. Bent and coiled tails were the most frequent flagellar abnormalities and were observed in most samples. In most fertile men, abnormal sperm cells generally had fewer than two associated abnormalities (95th percentile of MAI = 1.92).

Comparison of fertile men, infertile men and TGCT patients

The distribution profiles of the percentages of normal spermatozoa, all morphological sperm abnormalities and MAI, in fertile men, infertile men and TGCT patients, are presented in Fig. 2. The distribution of the percentage of morphologically normal spermatozoa was essentially Gaussian in the three populations, but with a sharp shift towards lower values in infertile men and TGCT patients. More than 50% of TGCT patients and almost 40% of infertile men had a percentage of morphologically normal spermatozoa below the 5th percentile for fertile men.

The frequency of each morphological abnormality was significantly higher in infertile men and TGCT patients than in fertile men. In the three groups of men studied, the distributions of the incidence of abnormal post-acrosomal regions, abnormal acrosomal regions, bent or misaligned midpieces, coiled tails and of the MAI were also almost Gaussian, but the curves were displaced to the right for infertile men and TGCT patients in comparison with the curve for fertile men. Other categories of sperm abnormalities followed an exponential decay distribution profile that was generally less steep in infertile men and TGCT patients than in fertile men. More than a quarter of TGCT patients had values above the 95th percentile for fertile men, for the categories microcephalic sperm cells, abnormal post-acrosomal and acrosomal regions, ERC and short tail. Similarly, more than a quarter of the male partners from infertile couples had a frequency of abnormalities greater than that of fertile men for the categories abnormal post-acrosomal and acrosomal regions, thin midpiece, bent or misaligned midpiece and short tail. In addition, some sperm defects, microcephalic sperm cells, macrocephalic sperm cells, abnormal acrosomal regions and ERC, were significantly more frequent in TGCT patients than in infertile men. Interestingly, in comparisons restricted to normozoospermic men, significant differences were still found between the three groups studied, for most of abnormality categories (Table II).

The usefulness of the proposed three levels of abnormality

The broad overlap between the distribution profiles of the various morphological abnormalities in fertile and infertile men precluded the

### Table I: Distributions of morphologically normal spermatozoa (including borderline cells) and categories of spermatozoa with head, midpiece and principal piece abnormalities, in 926 fertile men.

<table>
<thead>
<tr>
<th>Abnormality Category</th>
<th>Distribution range (percentiles)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st</td>
</tr>
<tr>
<td>Normal spermatozoa</td>
<td></td>
</tr>
<tr>
<td>Tapered head</td>
<td>7</td>
</tr>
<tr>
<td>Thin head</td>
<td>21</td>
</tr>
<tr>
<td>Microcephalic</td>
<td>11</td>
</tr>
<tr>
<td>Macrocephalic</td>
<td>3</td>
</tr>
<tr>
<td>Multiple heads</td>
<td>4</td>
</tr>
<tr>
<td>Abnormal post-acrosomal region</td>
<td>52</td>
</tr>
<tr>
<td>Abnormal acrosomal region</td>
<td>74</td>
</tr>
<tr>
<td>Excess residual cytoplasm</td>
<td>8</td>
</tr>
<tr>
<td>Thin midpiece</td>
<td>1</td>
</tr>
<tr>
<td>Bent or misaligned tail</td>
<td>16</td>
</tr>
<tr>
<td>No tail</td>
<td>9</td>
</tr>
<tr>
<td>Short tail</td>
<td>4</td>
</tr>
<tr>
<td>Irregularly shaped tail</td>
<td>5</td>
</tr>
<tr>
<td>Coiled tail</td>
<td>26</td>
</tr>
<tr>
<td>Multiple tails</td>
<td>3</td>
</tr>
<tr>
<td>Multiple Anomalies Index (MAI)</td>
<td>2.08</td>
</tr>
</tbody>
</table>

*See Fig. I for definitions.
Figure 2. Curves best fitting the distributions of the various sperm morphology characteristics assessed in fertile male partners of pregnant women (black line), male partners from infertile couples (blue line) and TGCT patients (red line). The vertical line represents the 95th percentile threshold in fertile men (5th percentile for the percentage of morphologically normal spermatozoa), and the percentages in blue and red represent the percentages for infertile men and TGCT patients, respectively, above the 95th percentile threshold (below the 5th percentile threshold for the percentage of normal spermatozoa). (A) Head abnormalities, (B) midpiece and principal piece abnormalities, (C) Multiple Anomalies Index (MAI) and normal spermatozoa.
Figure 2 Continued.
definition of a single meaningful threshold for each abnormality category. We found that the experience-based threshold chosen for the definition of the HLA category (the 99th percentile in the population of infertile men) differed considerably between abnormalities. For example, it was 95% for abnormal acrosomal regions, but only 5% for macrocephalic spermatozoa and 4% for a thin midpiece (Table III). Similarly, the threshold defining LLA (the 95th percentile for the population of fertile men) was ≤3% for seven morphological abnormalities (tapered head, macrocephalic sperm, multiple heads, thin midpiece, short tail, irregularly shaped tail and multiple tails), but as high as 60% for abnormal acrosomal regions (Table III).

Figure 3 illustrates the possible relevance of the three levels of abnormality proposed for diagnosis or research purposes. It shows the sperm morphology profiles for six men, one of whom was fertile, the other five being infertile with several types of characterized teratozoospermia of known genetic origin. The graphical representation of the abnormal morphology profile shows both the exact level of each category of abnormality and the relative level according to the defined three-level/color ranges of abnormalities taken from the thresholds defined in the distributions in large groups of fertile and infertile men. This figure demonstrates that, in such clinical situations, the simple recording of the percentage of normal spermatozoa or the determination of MAI is not informative for diagnosis (see Fig 3A, B and D). In contrast, the different categories of abnormality level—low (LLA), high (HLA) or intermediate (ILA)—based on the frequencies determined according to our standardized definitions for each category of abnormality, provided a clear picture of the morphological background noise (LLA for all abnormalities) characteristic of fertile men, or of the combination of high levels of head or midpiece or principal piece defects characterizing particular pathological morphologies. Examples C and D refer to well-known genetic ‘monomorphic syndromes’ (note the ‘normal’ MAI) that are easily discernible. Examples E and F show more complex profiles (several yellow and red categories and a high MAI) that deserve investigations. A recent study has shown the genetic origin of the combination of the various tail defects; here we show that noticeable levels of defects of the head or midpiece compartment may be associated with tail defects suggesting a possible different genetic origin. In addition, examples in Fig. 3 show that: (i) a relatively high percentage of certain abnormalities may correspond to LLA and therefore has little meaning as for the abnormality ‘Abnormal post-acrosomal region’ (~25–30%, see A, D, B); (ii) a relatively low percentage of certain abnormalities may correspond to HLA and therefore has an important meaning, for example ‘Thin midpiece’ or ‘Irregular tail’ (B, for both) and (iii) some semen samples may have a rate of morphologically normal spermatozoa in the normal limits but may have a HLA for some categories of anomalies (D and B). These examples provide an indication of the potential usefulness of this approach for future studies aiming to identify similar or new specific profiles related to genetic or environmental factors.

**Discussion**

This study is the first to report the distribution of various head, midpiece and principal piece defects encountered in human spermatozoa in large groups of fertile men, infertile men and patients with a testicular cancer, through the use of a standardized classification method. For all the characteristics studied, the distributions observed for infertile men and TGCT patients deviated from that for fertile men, with several abnormality categories even displaying significantly higher frequencies in TGCT patients than in infertile men.

**Information obtained from a standardized detailed sperm morphology assessment**

The incidence of tapered heads was low in all three groups and lower than published values (Andrade-Rocha, 2007), possibly due to the lack of a precise definition of tapered heads in other studies (thin heads or amorphous head shapes may have been included). Thin heads (normal sperm head length but narrower width) have rarely been described, probably owing to confusion with tapered heads. Thin heads were more frequently observed in infertile patients and men with TGCT and may reflect abnormal nuclear morphogenesis (Auger et al., 1990) and fragmented DNA (Gandini et al., 2000). The relatively high incidence of this abnormality in fertile men is probably dependent on complex, multiple, environmental factors rather than genetic determinants. A high incidence of thin heads has been reported in varicocele (Prasivoravong et al., 2014) and experimental cryptorchidism (Mieusset et al., 1987).
Table II Comparison of men with a normal total sperm count and motility according to World Health Organization (2010) criteria, for the various standardized sperm morphological categories.

<table>
<thead>
<tr>
<th></th>
<th>Fertile men</th>
<th>Infertile men</th>
<th>TGCT patients</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 859</td>
<td>n = 930</td>
<td>n = 127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal spermatozoa</td>
<td>51.3 ± 14.4 SD</td>
<td>36.4 ± 16.4*</td>
<td>28.6 ± 13.2***</td>
<td>291</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53 (41–62)†</td>
<td>37 (24–48)</td>
<td>28 (20–38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tapered head</td>
<td>0.7 ± 1.5</td>
<td>1.2 ± 2.6*</td>
<td>1.5 ± 2.0*</td>
<td>18</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>0 (0–1)</td>
<td>0 (0–1)</td>
<td>1 (0–2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thin head</td>
<td>4.5 ± 5.1</td>
<td>7.4 ± 8.0*</td>
<td>8.0 ± 6.2*</td>
<td>55</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>3 (1–7)</td>
<td>5 (2–10)</td>
<td>7 (3–11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microcephalic</td>
<td>2.4 ± 2.6</td>
<td>4.5 ± 5.7*</td>
<td>8.2 ± 7.3***</td>
<td>89</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>2 (1–3)</td>
<td>3 (1–6)</td>
<td>6 (2–13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macrocephalic</td>
<td>0.3 ± 0.8</td>
<td>0.4 ± 0.9</td>
<td>0.7 ± 1.1***</td>
<td>8</td>
<td>&lt;0.0005</td>
</tr>
<tr>
<td></td>
<td>0 (0–0)</td>
<td>0 (0–0)</td>
<td>0 (0–1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple heads</td>
<td>0.4 ± 1.0</td>
<td>0.7 ± 1.8*</td>
<td>0.6 ± 1.1</td>
<td>7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Abnormal post-acrosomal region</td>
<td>20.5 ± 11.2</td>
<td>29.8 ± 16.2*</td>
<td>31.8 ± 17.1*</td>
<td>115</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53 (41–62)†</td>
<td>28 (17–40)</td>
<td>29 (20–41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal acrosomal region</td>
<td>32.5 ± 14.6</td>
<td>42.4 ± 18.1*</td>
<td>55.6 ± 16.0**</td>
<td>165</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Excess residual cytoplasm</td>
<td>1.0 ± 1.7</td>
<td>2.5 ± 3.2*</td>
<td>3.3 ± 3.1***</td>
<td>105</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Thin midpiece</td>
<td>0.1 ± 0.3</td>
<td>0.3 ± 0.6*</td>
<td>0.3 ± 0.8*</td>
<td>31</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Bent or misaligned tail</td>
<td>5.8 ± 3.7</td>
<td>9.4 ± 5.3*</td>
<td>7.4 ± 4.9***</td>
<td>140</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>5 (3–8)</td>
<td>8 (6–12)</td>
<td>6 (4–11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No tail</td>
<td>1.6 ± 2.1</td>
<td>1.7 ± 2.0</td>
<td>1.7 ± 2.0</td>
<td>1</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>1 (0–2)</td>
<td>1 (0–2)</td>
<td>1 (0–2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short tail</td>
<td>0.3 ± 0.8</td>
<td>1.0 ± 1.7*</td>
<td>1.3 ± 3.0*</td>
<td>68</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Irregularly shaped tail</td>
<td>0.5 ± 1.1</td>
<td>1.2 ± 1.8*</td>
<td>0.9 ± 1.7*</td>
<td>51</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Coiled tail</td>
<td>6.3 ± 5.6</td>
<td>8.1 ± 5.9*</td>
<td>8.1 ± 7.0*</td>
<td>21</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>5 (3–8)</td>
<td>7 (4–11)</td>
<td>6 (3–11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple tails</td>
<td>0.4 ± 0.7</td>
<td>0.7 ± 1.0*</td>
<td>0.7 ± 0.8*</td>
<td>32</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Multiple Anomalies Index (MAI)</td>
<td>1.58 ± 0.20</td>
<td>1.72 ± 0.24*</td>
<td>1.81 ± 0.24***</td>
<td>127</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>1.57 (1.43–1.71)</td>
<td>1.71 (1.55–1.86)</td>
<td>1.78 (1.62–2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

†Mean ± SD.
‡Median (IQ range).

For each morphological category, differences between the three groups of men were investigated by one-way analysis of variance taking unequal variances into account (Brown–Forsythe test). For pair-wise comparisons, post hoc Tukey tests were carried out, with: *P < 0.01 in comparisons with fertile men, **P < 0.01 in comparisons with infertile men.
(Russell et al., 1989). However, it remains unclear whether and how this disease interferes with sperm morphogenesis. Most infertile men had an incidence of ERC below 20%. Only a few cases had a markedly higher frequency of ERC, from 30 to 44%, probably corresponding to a highly specific context or causal factor. In these extreme cases, the high incidence of ERC below 20%.

Abnormal sperm morphology

1. Multiple heads
2. Abnormal post-acrosomal region
3. Abnormal acrosomal region
4. Excess residual cytoplasm
5. Thin midpiece
6. Bent or misaligned midpiece
7. No tail
8. Short tail
9. Irregularly shaped tail
10. Coiled tail
11. Multiple tails

According to our definition, coiled tails category includes coiling with the head inside or outside the coils, two forms of coiling with different origins (Yeung et al., 2009). The incidence of all coiled tails in the infertile population, with a mean value of 10%, was similar to that reported by Yeung et al. (3% head-in-coil and 8% other coiled sperm) for a comparable population and higher than in the fertile men studied (mean = 6.5%). The percentage of men with more than 20% of spermatozoa with coiled tails was 10% for infertile men and 11% for TGCT patients, an incidence similar to that reported by Yeung and co-workers. This abnormality is positively associated with heavy smoking, and may fluctuate with levels of neutral α-glucosidase, suggesting the negative influence of a deleterious epididymal environment (Yeung et al., 2009). The incidence of multiple tails was very low in all three populations, typically below 1%, the maximum incidence in this series being 8%. A higher incidence, which may affect the vast majority of the spermatozoa, is rare and associated with macrocephalic heads in the macrocephalic sperm syndrome caused by homozygous mutations of the aurora kinase C gene (Dieterich et al., 2007).

Clinical relevance of a standardized and detailed assessment of normal and abnormal sperm morphology

The 5th percentile value of normal spermatozoa in our population of fertile men, taken as the reference threshold, as in WHO5, was 23%. The similarity in the groups of fertile men studied here and in the study by Cooper et al. (2010) made it possible to compare our 23% threshold with the 4% threshold obtained using the WHOS criteria that consider all encoding an inner arm heavy chain dynein (Ben khelifa et al., 2014).
Figure 3 Examples of six profiles of morphological sperm abnormalities with the corresponding percentage of normal spermatozoa and MAI, based on the standardized classification, with a representation of low (green), intermediate (yellow) and high (red) levels of abnormality, based on the distributions in fertile and infertile men (see Statistical methods). (A) A fertile man with low levels of abnormality; (B) an infertile patient with a normal frequency of normal spermatozoa and MAI despite an abnormally high level of double-headed spermatozoa; (C) an infertile patient with a globozoospermia-like profile (Gui-chaoua et al., 2009) and no normal sperm cell, and a relatively low MAI due principally to very high levels of microcephalic sperm cells, essentially combined with an abnormal acrosomal region; (D) an infertile patient with fibrous sheath dysplasia (Chemes et al., 1987), despite normal levels for the percentage of normal spermatozoa and MAI; (E) an infertile patient with high levels of multiple morphological abnormalities of the flagella (Ben Khelifa et al., 2014) and an exceptionally high level of thin midpieces, resulting in low numbers of normal spermatozoa and a high MAI; (F) an infertile patient with MMAF together with high levels of thin and microcephalic heads resulting in a very high MAI. ERC, excess residual cytoplasm; DFS, dysplasia of the fibrous sheath; MMAF, multiple morphological abnormalities of the flagella.
Abnormal sperm morphology

The marked difference between these two values could correspond primarily to the inclusion or exclusion of the borderline aspects in the normal sperm categorization (Eliasson, 2010; Handelsman and Cooper, 2010). The use of a set of criteria for borderline aspects (Rothmann et al., 2013) could be useful in answering the question of whether the borderline aspects per se matter or not.

A recent epidemiological study (Buck Louis et al., 2014) showed that normal sperm morphology, whether according to the traditional WHO approach (including borderline forms) or to the 'strict criteria' (excluding borderline forms, as in WHOS), displays a similar level of association with time to pregnancy. However, a possible bias in this study cannot be ruled out, because a given laboratory does not generally use both approaches during routine semen analysis. Using the 5th percentile value of normal spermatozoa in fertile men as a threshold, more than one-third of unselected infertile men and half of TGCT patients were below this threshold. No such discrimination was reported when using the 'strict criteria' classification, for comparisons of fertile men and selected infertile men (Guzick et al., 2001) or with use of the reported 4% threshold for morphologically normal sperm cells according to the pregnancy rate in IUI (Deveneau et al., 2014).

The incidence of various sperm abnormalities or the resulting MAI may be associated with the probability of pregnancy (Jouannet et al., 1988; Slama et al., 2002; Buck-Louis et al., 2014). However, this does not imply that a single category of abnormalities considered independently of other co-variables could serve as a powerful fertility prognostic tool. Thus, none of the abnormality categories studied can be used in isolation, to discriminate, with high specificity and sensitivity, between the male partners of pregnant women, and infertile men or testicular cancer patients, as previously reported for other semen characteristics (Ombelet et al., 1997; Björndahl, 2011). However, high frequencies of some crucial categories of abnormalities may reflect overall changes to the spermiogenetic process, with functional consequences resulting in a lower probability of natural conception within a short period of time (Jouannet et al., 1988), or a lower rate of fertilization (Jeulin et al., 1986). This should be borne in mind in the management of infertile couples. Is it reasonable to determine the profile of all morphological abnormalities systematically in ART? Indeed, the standardized assessment of sperm abnormalities requires attention to detail, rather than a simple dichotomous classification of normal and abnormal spermatozoa. Training and quality control are required (Eustache and Auger, 2003). In our experience, acceptable levels of variability can be achieved after an initial period of theoretical and practical training no longer than that for the evaluation of sperm motility. Quality controls (e.g. in Supplementary data, Figure S1) indicate non-critical variations provided that regular quality controls ensure the maintenance of competence (the episodic discrepancies observed are no greater for the assessment of sperm defects than for normal sperm assessments). A detailed standardized assessment of morphology abnormalities could be recommended for the first andrology check-up for infertile couples, regardless of whether ART is to be carried out. It may be useful if carried out in addition to an andrological check-up in which medical history, lifestyle factors and work environment are considered. In contrast, owing to the known stability of sperm morphological features, a detailed morphological assessment is not required in the subsequent follow-up or during ART. However, if a factor (fever, medication, temporary exposure to toxicants etc.) is thought to have influenced the sperm morphological profile, a repeat assessment should be considered.

Towards a new quantification of sperm abnormalities for clinical and research purposes

We propose three levels of defects based on the percentiles for fertile and infertile men: a low level ≤ 95th percentile for fertile men; an intermediate level > 95th percentile for fertile men and < 99th percentile for infertile men and a high level ≥ 99th percentile for infertile men. This three-level stratification of morphological abnormalities is potentially useful for studies considering genetic determinants or the role of environmental factors capable of disrupting sperm morphogenesis, as illustrated in Fig. 3. The low level of abnormality essentially corresponds to non-specific background noise. In contrast, the HLA, which excludes most fertile men and corresponds to extreme values in the infertile population, may be considered a biomarker of a causal genetic or environmental factor requiring investigation. Despite, the large overlap between fertile and infertile men with intermediate levels of abnormality, the intermediate category should not be considered as a random signal in any of the male populations studied (see Fig. 3E and F).

The potential of a standardized assessment of sperm defects for epidemiological, genetic and basic studies

A few studies have reported associations between the extent of various sperm defects and pathological conditions, such as varicocele or cryptorchidism (Auger et al., 2001; Andrade-Rocha, 2007; Cakan et al., 2008), environmental or occupational exposures (Bigelow et al., 1998; Gebreegziabher et al., 2004; Hansen et al., 2010). However, the imprecise definition of sperm defects in most of these studies constitutes a confounding factor. An interesting alternative approach for basic or epidemiological studies is the combination of light microscopy with computer vision, which is highly accurate and precise, provided that pre-analytical and analytical set-ups are optimized (Bellastella et al., 2010). Several reports have demonstrated the relevance of this approach for such studies (Schrader et al., 1990; Fenster et al., 1997). The measurement of various morphometric features of all sperm cells and the visual estimation of well-characterized abnormalities may be complementary approaches as recently reported (Buck Louis et al., 2014). However, it should be pointed out that computer vision is mainly applied for the measurement of a set of features mainly of the sperm head compartment, and not the midpiece and principal piece. In contrast, the visual approach, although not allowing precise and reproducible measurements of the various sperm compartments, is more appropriate than computer vision for pattern recognition of an entire sperm cell. Overall, the choice of the method used depends on the laboratory equipment (only a minority of reproductive laboratories have analytical cytometry facilities) and of the aims of the study (Auger, 2010, for a detailed comparison of the advantages and limits of visual and computer-aided approaches). Only rare morphological syndromes such as globozoospermia, macrocephalia, decapitated sperm syndrome and fibrous sheath dysplasia have been described in men consulting for infertility, and several causal mutations have been identified (Coutton et al., 2015). However, these mutations do not account for all cases and the variability of clinical forms observed suggests that other mutations or other genes may be involved. Partial forms of these syndromes in which apparently normal spermatozoa coexist with those of the morphological type of interest are less rarely
observed. More generally, particular combinations of abnormalities with high frequencies may reflect complex and variable genetic disorders. For this reason, a detailed, standardized analysis of the sperm defects, accounting for the respective percentages of each defect, may help to distinguish between phenotypes.

In conclusion, the sole assessment of the percentage of morphologically normal spermatozoa, as usually performed in infertile couples before ART, and also in epidemiological studies, conceals the differential effects of exposures or genetic determinants on head or tail compartments and substructures, and on sperm fertilizing ability. Based on our results showing noticeably different distributions of the sperm abnormalities in fertile men, infertile men and testicular cancer patients, a standardized assessment of morphological abnormalities of spermatozoa including the use of a trichotomous categorization of abnormalities based on their level of occurrence is proposed. It would be useful at a first andrological check-up rather than during the routine management of infertile couples by ART. This approach cannot provide information about possible functional consequences and it is clearly not dedicated to deciphering the underlying modes of action of various pathological or environmental conditions on the sperm morphogenesis process. However, by improving sperm phenotyping, it constitutes a useful step forward for clinical, genetic and epidemiological studies.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Authors’ roles

J.A. and F.E. designed the study, verified all data collected and carried out the data analysis. J.A., F.E. and P.J. interpreted the data, wrote the article and prepared the tables and figures.

Funding

Part of this study was supported by contract BMH4-CT96-0314 from the European Union.

Conflict of interest

None declared.

References


Cooper TG. Cytoplasmic droplets: the good, the bad or just confusing? Hum Reprod 2005;20:9 – 11.


Abnormal sperm morphology


