



UNIVERSIDAD DE GRANADA

PROGRAMA DE DOCTORADO EN BIOMEDICINA

TESIS DOCTORAL

**AVANCES EN EL CONTROL DE CALIDAD
EXTERNO EN EL LABORATORIO DE
EMBRIOLOGÍA CLÍNICA**

Luis Martínez Granados

Granada, 2018

Editor: Universidad de Granada. Tesis Doctorales
Autor: Luis Martínez Granados
ISBN: 978-84-9163-813-1
URI: <http://hdl.handle.net/10481/50141>

D. JOSÉ ANTONIO CASTILLA ALCALÁ, PROFESOR ASOCIADO LABORAL DEL DEPARTAMENTO DE ANATOMÍA Y EMBRIOLOGÍA HUMANA DE LA UNIVERSIDAD DE GRANADA. FACULTATIVO ESPECIALISTA DE ÁREA DE LA UNIDAD DE GESTIÓN CLÍNICA DE LABORATORIO CLÍNICO Y UNIDAD DE GESTIÓN CLÍNICA DE OBSTETRICIA Y GINECOLOGÍA DEL HOSPITAL UNIVERSITARIO VIRGEN DE LAS NIEVES DE GRANADA.

CERTIFICO:

Que D. Luis Martínez Granados ha realizado bajo mi dirección el trabajo de Tesis Doctoral sobre el tema: “Avances en el control de calidad externo en el laboratorio de embriología clínica”. Una vez finalizado este trabajo, ha sido revisado y estoy conforme con su presentación para obtener el grado de Doctor, siempre que así lo considere la Universidad de Granada.

Granada, 22 de diciembre de 2017

Fdo: José Antonio Castilla Alcalá

El doctorando, D. Luis Martínez Granados y el director de tesis, D. José Antonio Castilla Alcalá, garantizamos que al firmar esta tesis doctoral el trabajo ha sido realizado por el doctorando bajo la dirección del director de tesis y hasta donde nuestro conocimiento alcanza en la realización del trabajo se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 22 de diciembre de 2017

Director de tesis

Doctorando

Fdo; José Antonio Castilla Alcalá

Fdo: Luis Martínez Granados

Agradecimientos

Al Dr. José Antonio Castilla, por haberme permitido desarrollar, bajo su dirección, este proyecto y por inculcarme la importancia de la calidad, especialmente en el laboratorio de reproducción humana asistida. Me siento realmente afortunado por haber recibido sus consejos y su orientación en este oficio.

A la Dra. M^a Carmen Gonzalvo, por su implicación en mi formación como embriólogo clínico. Por enseñarme hasta el mínimo detalle del laboratorio. Ha sido un regalo poder disfrutar de su labor docente.

A la Dra. Ana Clavero, por haberme hecho participe de los proyectos de los que ha sido responsable. Gracias por formar parte de mi trayectoria.

A mis compañeros de laboratorio, por su paciencia, cuando han debido enseñarme y cuando han debido sufrirme.

A Marisa, por transmitirme la pasión por este oficio, por darme siempre el “otro punto de vista”, por sus consejos que me han moldeado como embriólogo. Por enseñarme que el silencio que precede al susto en las películas de miedo puede ser muy maravilloso. Gracias por esto y por todo lo demás.

A Puri, por ser una compañera de laboratorio inmejorable. Por sus infinitas ganas de compartir sus conocimientos, de las que tanto me he aprovechado y sin duda, por guiarme en todo el tema burocrático que conlleva depositar y defender una tesis doctoral. Gracias por tu amistad y por tantos momentos inolvidables, dentro y fuera del laboratorio.

A todos los miembros de la Unidad de Reproducción del Hospital Universitario Virgen de las Nieves, por su atención y amabilidad siempre que he necesitado algo de ellos.

A los miembros del Grupo de interés de Calidad de ASEBIR, especialmente a Nereyda Ortiz, María Serrano y Antonio González Utor por su implicación e interés en este trabajo. Por sus aportaciones y críticas, siempre constructivas. Si he visto más lejos, es porque estoy sentado sobre hombros de gigantes.

A todos los profesionales de los laboratorios participantes en el programa de control de calidad externo para el laboratorio de reproducción, sin cuya participación, esta tesis no hubiese sido posible.

A mis padres, por haberme animado a seguir formándome y por su apoyo en toda mi formación como embriólogo. Gracias por tantos consejos que antes o después he tenido que reconocer como buenos y necesarios, por su cariño, por las conversaciones interminables y, en definitiva, gracias por creer en mí.

A mi hermano

El aumento del conocimiento
depende por completo de
la existencia del desacuerdo.

Karl Popper (1902-1994)

Los resultados de esta Tesis Doctoral han sido publicados en las siguientes revistas científicas:

- Martínez-Granados L, Serrano M, González-Utor A, Ortíz N, Badajoz V, Olaya E, Prados N, Boada M, Castilla JA. Special Interest Group in Quality of ASEBIR (Spanish Society for the Study of Reproductive Biology). Inter-laboratory agreement on embryo classification and clinical decision: Conventional morphological assessment vs. time lapse. PLoS One. 2017;12:e0183328. Factor de impacto: 3.11 (Q1).
- Martínez-Granados L, Serrano M, González-Utor A, Ortíz N, Badajoz V, López-Regalado ML, Boada M, Castilla JA. Special Interest Group in Quality of ASEBIR (Spanish Society for the Study of Reproductive Biology). Reliability and agreement on embryo assessment: five years of an external quality control programme. Reprod Biomed Online 2018. *In press*. Factor de impacto: 2.59 (Q1).

Y han sido presentados en las siguientes reuniones y congresos de sociedades científicas nacionales:

- VIII Congreso Nacional de la Asociación para el Estudio de la Biología de la Reproducción (ASEBIR). (San Sebastián, España) Noviembre 2015.
- XXIII Reunión Científica de la Sociedad Andaluza de Análisis Clínicos. (Almería, España) Marzo 2016.
- 31º Congreso Nacional de la Sociedad Española de Fertilidad. (Málaga, España) Mayo 2016.
- IX Congreso Nacional de la Asociación para el Estudio de la Biología de la Reproducción (ASEBIR). (Madrid, España) Noviembre 2017.

Índice

Resumen	1
Introducción	21
1. Técnicas de reproducción asistida.....	23
1.1. Inseminación artificial	23
1.2. FIV/ICSI	24
1.3. Otras técnicas.....	26
1.4. Vitricación ovocitaria y embrionaria	28
2. Evaluación morfológica clásica de ovocitos y embriones	29
2.1. Evaluación ovocitaria.....	29
2.2. Evaluación del cigoto.....	31
2.3. Evaluación embrionaria	34
2.4. Evaluación de blastocistos	40
2.5. Sistemas de catalogación embrionaria	41
3. Nuevas técnicas de evaluación embrionaria: Equipos <i>time-lapse</i>	46
3.1. Equipos <i>time-lapse</i>	46
3.2. Ventajas y limitaciones de los equipos <i>time-lapse</i>	47
3.3. Parámetros cinéticos embrionarios.....	50
4. Reproducibilidad de la evaluación embrionaria	56
5. Calidad en el laboratorio de reproducción humana asistida	60
5.1. Calidad: Generalidades	60
5.2. Gestión y garantía de la calidad	61
5.2.1. Gestión de la Calidad	61
5.2.2. Garantía de la Calidad.....	61
5.3. Sistema de calidad y laboratorio.....	62
5.4. Sistema de calidad y laboratorio de reproducción humana asistida	65
5.5. Procesos del laboratorio de reproducción humana asistida	68
5.6. Calidad de las fases de las actividades de procedimiento.....	71
5.6.1. Calidad de la fase pretécnica	71

5.6.2. Calidad de la fase técnica.....	79
5.6.3. Calidad de la fase posttécnica	86
5.7. Control de calidad	88
5.7.1. Control de calidad interno	88
5.7.2. Control de calidad externo	92
5.8. Programa de control de calidad externo en el laboratorio de reproducción humana asistida	97
5.8.1. Control de calidad externo del análisis de semen.....	100
5.8.2. Control de calidad externo de evaluación embrionaria.	101
5.8.3. Control de calidad externo de evaluación embrionaria con equipos <i>time-lapse</i>	103
5.8.4. Control de calidad externo para la realización de ensayos de citotoxicidad.....	104
6. Interpretación de resultados de estudios sobre reproducibilidad	106
6.1. Fiabilidad y acuerdo	106
6.2. Guía para Publicar Estudios de Fiabilidad y Acuerdo (GRRAS)	109
6.3. Estadísticos recomendados en estudios de fiabilidad y acuerdo	111
6.3.1. Parámetros de fiabilidad para variables cualitativas (nominales, dicotómicas y ordinales)	112
6.3.2. Parámetros de fiabilidad para variables cuantitativas	114
6.3.3. Parámetros de acuerdo para variables cualitativas	116
6.3.4. Parámetros de acuerdo para variables continuas.....	117
Hipótesis y justificación	119
1. Hipótesis	121
2. Justificación.....	121
Objetivos	123
Material y métodos	127
1. Acuerdo y fiabilidad entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica	129
1.1. Participación	129
1.2. Videos	129
1.3. Visualización de las imágenes	130
1.4. Boletín de resultados	130

1.5. Evaluación de los videos	132
2. Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante <i>time-lapse</i>	134
2.1. Participación	134
2.2. Videos.....	135
2.3. Esquema organizativo del ensayo piloto de control de calidad externo de evaluación morfológica embrionaria utilizando equipos <i>time-lapse</i>	136
2.4. Boletín de resultados.....	136
3. Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad	138
3.1. Participación	138
3.2. Esquema organizativo del control de calidad externo de citotoxicidad ...	139
3.3. Productos y reactivos para el chequeo por bioensayos.....	139
3.3.1. Productos	139
3.3.2. Reactivos.....	139
3.3.3. Envios.....	140
3.4. Tratamiento de los productos a analizar.....	141
3.5. Boletín de resultados.....	141
4. Métodos estadísticos	142
4.1. Fiabilidad y acuerdo entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica	142
4.1.1. Fiabilidad.....	142
4.1.2. Acuerdo.....	143
4.1.3. Número de observaciones.....	143
4.2. Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante <i>time-lapse</i>	144
4.3. Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad.....	145
Resultados	147
1. Acuerdo y fiabilidad entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica.....	149

1.1. Catalogación embrionaria	149
1.2. Características morfológicas embrionarias	152
1.3. Destino del embrión	163
2. Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante equipos <i>time-lapse</i>	165
2.1. Tiempos del desarrollo embrionaria en los equipos <i>time-lapse</i>	165
2.2. Catalogación embrionaria	170
2.3. Características morfológicas embrionarias	174
2.4. Destino del embrión	176
3. Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad.....	178
3.1. Influencia de la experiencia de los laboratorios en la participación en un programa de control de calidad externo de citotoxicidad	183
3.2. Correlación entre el grado de acuerdo en medos de cultivo y puntas de pipeta en el test de citotoxicidad	187
Discusión	189
1. Acuerdo y fiabilidad entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica	192
2. Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante <i>time-lapse</i>	199
3. Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad.....	206
Conclusiones	211
Bibliografía	215

Siglas y abreviaturas

AAB: Asociación americana de bioanalistas

AEFA: Asociación española de farmacéuticos analistas

AENOR: Asociación española de normalización y certificación

AMFE: Análisis modal de fallos y efectos

ASEBIR: Asociación para el estudio de la biología de la reproducción

CCExt: Control de calidad externo

CCI: Coeficiente de correlación intraclase

CCInt: Control de calidad interno

CEIFER: Centro de estudio e investigación de la fertilidad

CT: Citotoxicidad

CV: Coeficiente de variación

DS: Desviación estándar

DT: División temprana

EME: Evaluación morfológica embrionaria

EMEC: Evaluación morfológica embrionaria clásica

ESHRE: Sociedad europea de reproducción humana y embriología

FIV: Fecundación in-vitro

FN: Falsos negativos

FP: Falsos positivos

IAC: Inseminación artificial conyugal

IAD: Inseminación artificial de donante

ICSI: Inyección intracitoplasmática de espermatozoide

LRHA: Laboratorio de reproducción humana asistida

MCI: Masa celular interna

NC: No conformidades

NGS: *Next generation sequencing* (Secuenciación de siguiente generación)

OMS: Organización mundial de la salud

PGT: Test genético preimplantacional

PN: Pronúcleos

TRA: Técnicas de reproducción asistida

TSE: Test de supervivencia espermática

VPN: Valor predictivo negativo

VPP: Valor predictivo positivo

Resumen

Introducción

Las técnicas de reproducción asistida son fundamentales para atender los deseos reproductivos de las parejas estériles y de las mujeres sin pareja masculina, para evitar la transmisión de enfermedades infecciosas y hereditarias o para preservar la fertilidad.

El laboratorio de reproducción debe estar orientado a satisfacer las necesidades de sus usuarios mediante procesos de probada seguridad y eficiencia. Para ello es clave la implantación de sistemas de calidad y programas de control de calidad internos (CCInt) y externos (CCExt). Actualmente, diferentes sociedades científicas han recomendado la participación en programas de CCExt de evaluación morfológica embrionaria y de citotoxicidad (CT) (OMS, 2010; Hurtado de Mendoza et al., 2015; ESHRE Guideline Group on Good Practice in IVF Labs et al., 2016), ya que permiten conocer la fiabilidad entre laboratorios y su exactitud, en el primer caso, y aumentar su eficiencia, en el segundo.

1. Análisis de ovocitos y embriones.

La evaluación de la calidad embrionaria es clave en el laboratorio de embriología clínica, sin embargo, diversos factores pueden afectar a esta valoración, como las posibles diferencias intra e interobservador existentes en la interpretación de los criterios de clasificación de cigotos y embriones.

Algunos autores como Arce et al. (2006) indican que la variabilidad intraobservador para la clasificación embrionaria es relativamente baja. Respecto a la variabilidad interobservador, es importante que todos los miembros de un equipo sigan un mismo criterio para poder trabajar como una unidad y tomar decisiones reproducibles y válidas. Esto implica una estandarización en los sistemas de evaluación embrionaria, así como la formación continuada del equipo de embriólogos (Arce et al., 2006). En 2008, la Asociación para el Estudio de la Biología de la Reproducción (ASEBIR) publica un manual, en el que se estandarizaba por primera vez los criterios de evaluación y clasificación embrionaria (Ardoy et al., 2008).

Sin embargo, la evaluación morfológica clásica, la más extendida en los laboratorios, está basada en determinaciones estáticas, presentando tres grandes desventajas:

1º Está vinculada a momentos puntuales, con la consiguiente pérdida de información sobre el resto del desarrollo embrionario.

2º Inevitablemente perturba las condiciones óptimas de cultivo.

3º Tiene un carácter subjetivo, dependiente del observador (Arce et al., 2006)

Frente a esta evaluación morfológica clásica ha aparecido, en los últimos años, la tecnología *time-lapse*, que permite el seguimiento del desarrollo embrionario en tiempo real. Estas plataformas ofrecen la posibilidad de cultivar embriones en un entorno controlado, capturando imágenes a intervalos definidos. Lo que ofrece varias ventajas frente a la microscopía estática convencional.

1º Permite el mantenimiento de condiciones óptimas de cultivo durante todo el desarrollo embrionario.

2º Permite hacer un monitoreo continuo de los embriones.

3º Reducción de posibles errores en la clasificación de embriones, debido a la posibilidad de visionar las imágenes indefinidamente.

No obstante, los sistemas *time-lapse* también presentan algunas limitaciones, como la incapacidad de rotar los embriones en observación.

Al ser una tecnología de reciente implantación, aun encontramos inconsistencias en la terminología utilizada a la hora de nombrar las variables morfocinéticas. Kaser y Racowsky (2014) pusieron de relieve que el mismo evento del desarrollo embrionario era denominado de hasta cuatro maneras diferentes según los autores. Esta terminología debería ser estandarizada.

En los años en los que solo existían los criterios morfológicos clásicos, diferentes encuestas y sondeos realizados a centros de reproducción asistida en España (Castilla et al., 2010), sacaron a la luz la gran disparidad de criterios que usaban los diferentes laboratorios para la clasificación embrionaria. La ausencia de criterios comunes afectaba tanto a los parámetros a evaluar, como a los

puntos de corte entre categorías, lo que restaba validez a dichos criterios, aumentando la variabilidad entre observadores. Para disminuir esta variabilidad se propusieron diferentes alternativas como la asistencia a actividades formativas específicas (Ruiz de Assin et al., 2011) y la participación en programas de control de calidad externos e internos (Arce et al., 2006; Castilla et al., 2010). En 2007, comenzó a realizarse en España un programa de CCExt para laboratorios de reproducción basado en la evaluación de la morfología embrionaria de manera estática. En 2011, los criterios de clasificación embrionaria de ASEBIR se incorporan al programa de CCExt.

En estos programas nuestro grupo demostró la importancia de utilizar valores verdaderos basados en la opinión de expertos (Ruiz de Assin et al., 2009). No obstante, en estos estudios aún no se aplicaban los criterios estandarizados de morfología embrionaria de ASEBIR por su reciente publicación en aquellos años.

Con la aparición de la tecnología *time-lapse* y su posterior implantación en numerosos laboratorios, ha surgido el mismo problema de falta de consenso y validez. Esta falta de consenso entre laboratorios conllevaría problemas como la imposibilidad de generar estudios multicéntricos con una valoración común de la calidad embrionaria, la dificultad en la interpretación de informes clínicos de otros laboratorios o en la comparación de los datos bibliográficos.

2.Evaluación de fungibles

El fungible de laboratorio utilizado en reproducción asistida que vaya a estar en contacto con gametos o embriones no debería tener una influencia negativa en la viabilidad de estos, por lo que la identificación de fungibles que puedan ser citotóxicos es esencial para el óptimo rendimiento de los laboratorios de embriología (Lierman et al., 2007).

La realización de bioensayos para determinar la utilización o no de materiales y medios es fundamental para evitar una disminución de los resultados o desechar materiales y/o medios no tóxicos, lo que implicaría costes extra. Un buen sistema de bioensayos de materiales y medios llevaría consigo una mejora de resultados con un menor coste.

Los programas de CCExt de CT en los laboratorios de reproducción están implantados desde hace tiempo, pero es necesario su estandarización ya que los fungibles que se evalúan son muy diversos (puntas de pipeta, placas de cultivo, catéteres de transferencia, medios de cultivo, etc.).

Tras lo comentado en la introducción, el objetivo de este trabajo es abordar diferentes aspectos de los programas de CCExt del laboratorio de embriología, utilizando una metodología estandarizada para realizar estudios de fiabilidad y acuerdo (GRRAS) (Kottner et al., 2011).

Hipótesis y justificación

1.Hipótesis

- La incorporación de una catalogación embrionaria estandarizada a un programa de CCExt de evaluación morfológica embrionaria clásica afectaría a la fiabilidad y acuerdo entre laboratorios y un panel de expertos.
- Por su reciente implantación, los equipos *time-lapse* carecen de controles de calidad externos, lo que podría restar validez a dichos sistemas al no controlarse la variabilidad entre observadores.
- Los diferentes fungibles utilizados en los CCExt de CT podrían influir en la capacidad de los laboratorios de detectar fungibles tóxicos.

2.Justificación

Uno de los retos a los que se enfrentan los laboratorios de reproducción asistida es el mantenimiento de la calidad en todos los procesos llevados a cabo en él. Para ello, se debe evaluar la eficacia de las decisiones y procedimientos realizados diariamente, identificar y corregir problemas, asegurar la certeza y la precisión de los procedimientos, y controlar la competencia del personal del laboratorio. Por este motivo es fundamental analizar y evaluar los procesos y técnicas que se llevan a cabo en el laboratorio de embriología de forma tradicional y aquellos procesos y técnicas nuevas que pretenden incorporarse a dichos laboratorios con objeto de asegurar y garantizar el buen funcionamiento de los mismos.

En los últimos años los dos avances más relevantes en el campo de la clasificación embrionaria han sido la publicación de los criterios estandarizados

de catalogación embrionaria de ASEBIR y la aparición de plataformas de morfocinética. Estas últimas se han instalado en nuestros laboratorios sin análisis previos de su reproducibilidad entre centros. Aunque actualmente se ha comprobado una baja variabilidad entre observadores del mismo centro, se desconoce si existen variaciones entre laboratorios a la hora de determinar los tiempos y las características del desarrollo embrionario.

Como se ha comentado anteriormente, es fundamental analizar y evaluar los procesos y técnicas que se llevan a cabo en el laboratorio de embriología. En la actualidad existen a nivel internacional distintos CCEExt para evaluar fungibles gametotóxicos, sin embargo, se desconoce si el rendimiento de un laboratorio en la evaluación de fungibles tóxicos depende del tipo de fungible evaluado.

Objetivos

1. Analizar la reproducibilidad entre laboratorios de embriología y un panel de expertos en evaluación morfológica embrionaria clásica.
2. Analizar el impacto de los equipos *time-lapse* en la variabilidad entre laboratorios de embriología en la evaluación morfológica embrionaria.
3. Valorar la influencia del tipo de producto a analizar en un programa de CCEExt de CT en el laboratorio de embriología.

Material y métodos

1. Acuerdo y fiabilidad entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica

Se analizan los datos recogidos en el programa de CCEExt de ASEBIR correspondientes a los años 2012-16, en los que participaron una media anual de 44 centros nacionales de forma voluntaria.

Se evalúan un total de 125 embriones a lo largo de los 5 años. En este programa, los centros acceden a una plataforma a través de internet en la que encuentran 3 proyectos con videos cigotos, embriones de día 2 y día 3 y 2 proyectos con videos de embriones de día 4 y día 5. La óptica y el aumento usado para la grabación fueron Hoffman y 400x, respectivamente. Las grabaciones se llevaron a cabo simulando la rutina diaria de valoración de la calidad embrionaria.

En este programa anual de CCEExt se evalúan, para los embriones en día 2 y día 3: número de células, asimetría de las blastómeras, porcentaje de fragmentación celular, presencia de multinucleación, presencia de vacuolas; y para embriones de día 5: tipo de blastocisto, masa celular interna y trofoectodermo (evaluadas según la catalogación de ASEBIR).

Además, los laboratorios deben clasificar cada embrión de día 2, día 3 y día 5 según la catalogación de ASEBIR. Todas las características analizadas, entre las 16-18 horas, 43-45 horas y 67-69 horas tras la microinyección espermática, permiten seguir un diagrama de flujo para cada embrión según ASEBIR, que indicará la calidad del embrión y por tanto su potencial de implantación. Clasificándose los embriones según ASEBIR en A (muy alta capacidad de implantación), B (alta capacidad de implantación), C (baja capacidad de implantación) y D (muy baja capacidad de implantación).

También deben decidir para los 3 primeros proyectos (embriones día 2 y día 3) que dos embriones debían ser transferidos y para los 2 últimos proyectos (embriones día 5) que embrión debía ser transferido. Para aquellos embriones que decidían no transferir, los centros debían decidir si los crioconservarían o los descartarían.

El grupo de interés de calidad de ASEBIR propone a un grupo de siete expertos. El grupo de expertos se reúne de forma presencial y anual para analizar las imágenes de la misma manera que lo hacen los centros.

Para cada variable analizada se considera, la respuesta dada por la mayoría de los centros y la respuesta consensuada por el grupo de expertos. La respuesta consensuada del grupo de expertos es utilizada como el valor verdadero. Aquellas características embrionarias en las que el grupo de expertos no llega a un consenso fueron eliminadas del estudio.

2.Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante *time-lapse*

Los centros participantes en este estudio fueron contactados mediante la secretaría técnica de ASEBIR. Esta llevó a cabo una invitación a todos sus asociados, vía e-mail, para la participación en un ensayo piloto de CCEExt. Indicando que la participación en este ensayo piloto sería voluntaria, por centro y

siendo condición indispensable disponer de alguna de las dos plataformas evaluadas (Embryoscope o PrimoVision).

A cada centro participante se le enviaron tres proyectos en un dispositivo USB. Cada proyecto incluía videos del desarrollo de entre 9-14 ovocitos hasta la fase de blastocisto. Estos videos se realizaron a partir de imágenes tomadas cada 5-20 minutos de 31 embriones para los usuarios de Embryoscope y 35 embriones para los usuarios de PrimoVision. Cada centro debía, en base a estos vídeos, clasificar los embriones en día 2, 3 y 5, siguiendo los criterios ASEBIR de clasificación embrionaria, y decidir, para cada proyecto, que embrión de día 5 debería ser transferido. Para aquellos embriones que decidían no transferir, los centros debían elegir si los crioconservarían o los descartarían.

Al mismo tiempo se facilitaron instrucciones con explicaciones sobre el funcionamiento del programa de control calidad y una base de datos para la recogida de tiempos y otras características morfológicas.

Se recogen los tiempos a los que ocurre: La extrusión del segundo corpúsculo polar (extrusión 2º CP), la aparición (aparición PN) y desaparición (desaparición PN) de los pronúcleos, cada una de las divisiones celulares (inicio de t2, t2, t3, t4, t5, t6, t7, t8, t9), tiempo en el que alcanza la fase de mórula (tmórula), tiempo en el que aparecen los primeros signos de blastulación (tsB), tiempo en el que alcanza la fase de blastocisto (tB), tiempo en el que comienza la expansión (tBE), tiempo en el que inicia la eclosión (tBHi).

Las características morfológicas recogidas fueron: Anomalías ovocitarias: Presencia de alguna característica morfológica anormal (agregados del REL, vacuolas, inclusiones, anomalías del espacio perivitelino, anomalías de la zona pelúcida, anomalías en el corpúsculo polar o en la forma del ovocito), número de pronúcleos, fragmentación celular, presencia de vacuolas, presencia de multinucleación y asimetría de las blastómeras.

3. Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad

Los datos analizados proceden del programa de CCExt de CT (años 2013-2016) organizado por ASEBIR.

En este programa se realiza un envío anual a los laboratorios participantes de dos fungibles (muestras de medios de cultivo embrionario y puntas de pipeta). Algunos de esos fungibles fueron tratados por la organización de tal manera que resultasen tóxicos. Los laboratorios participantes desconocen totalmente el tratamiento aplicado a cada fungible. Los fungibles habían sido previamente testados por la organización con el fin de confirmar su toxicidad o no toxicidad. Todos los centros utilizaron para valorar la toxicidad de los fungibles el test de supervivencia espermática y debían responder para cada fungible analizado si era tóxico o no tóxico.

4. Análisis estadístico

Para los cálculos anteriores, se utilizó el programa estadístico AgreeStat2015 [Advanced Analytics, LLC (USA)]. Se utilizó una ponderación cuadrática cuando fue necesaria aplicarla en algún parámetro. Para expresar la fuerza de concordancia para los estadísticos kappa y Gwet utilizamos los siguientes intervalos: 0 “no hay acuerdo”, 0,01-0,20 “leve”, 0,21-0,40 “justo”, 0,41-0,60 “moderado”, 0,61-0,80 “substancial” y 0,81-1 “casi perfecto” (Landis y Koch et al., 1977). Para una correcta interpretación del acuerdo ponderado se utilizaron los siguientes intervalos: 0.70-0.79 “pobre”; 0.80-0.89 “moderado”; 0.90-0.98 “fuerte”; 0.99-1.00 “casi perfecto” (Tworek et al., 2013).

Se consideró un alto grado de acuerdo cuando sobre un embrión o cigoto más del 75% de los laboratorios coincidían en su clasificación o decisión clínica.

Para la comparación de las diferentes variables cualitativas analizadas se utilizó el test de χ^2 con una significación del 5%. En caso de no cumplirse las condiciones de validez de dicho test, se utilizó el test de Fisher.

Los datos obtenidos en control de calidad externo de tecnología *time-lapse* usando los criterios estandarizados de ASEBIR se analizan con el paquete estadístico SPSS 19.0. Para cada tiempo analizado, se calculó el coeficiente de variación y el coeficiente de correlación intraclase, junto con el correspondiente intervalo de confianza al 95%. Los tiempos medios de cada evento de los embriones analizados con cada plataforma se compararon por el test de Student. Se utiliza el test de Chi cuadrado para comparar los datos categóricos. Todas las pruebas se realizaron con el nivel de significancia del 5%.

Se analizó la sensibilidad (S), especificidad (E), valor predictivo positivo (VPP), valor predictivo negativo (VPN) y acuerdo global de los bioensayos.

Se utilizó una representación del espacio ROC para interpretar los datos del test de supervivencia espermática (TSE).

Para determinar si existía relación entre el porcentaje de acierto al analizar medios de cultivo y el obtenido al analizar puntas de pipeta, se calculó el coeficiente de Correlación de Pearson.

Resultados

1. Acuerdo y fiabilidad entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica

Respecto a la evaluación morfológica embrionaria, el valor de fiabilidad obtenido entre laboratorios y el panel de expertos es sustancial (Gwet: 0.76; 95% IC: 0.70-0.84). El acuerdo entre laboratorios y el panel de expertos fue más alto en los embriones de buena calidad que en los de calidad más baja. Cuando se observa multinucleación o vacuolas, se obtienen bajos niveles de fiabilidad. En blastocistos, la característica que presenta las mayores discrepancias es la masa celular interna. En la decisión clínica sobre el embrión, la fiabilidad entre laboratorios y el panel de expertos es moderada (Gwet: 0.51; 95% IC: 0.41-0.60). Siendo más baja cuando el destino del embrión era crioconservarse, que cuando era transferirse o descartar.

2. Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante *time-lapse*

Para los dos equipos *time-lapse* estudiados (EmbryoScope™ y Primo Vision™) se observan dos periodos de alta variabilidad entre laboratorios en los tiempos de los eventos del desarrollo embrionario. Se observa un pico de variabilidad cuando los laboratorios deben determinar los primeros eventos del desarrollo (extrusión del segundo corpúsculo polar y aparición de los pronúcleos) y un segundo pico entre los tiempos correspondientes a 8 células y al estadio de mórula. En la mayoría de las variables cualitativas analizadas se observa una muy baja variabilidad entre centros, tanto en la evaluación morfológica clásica como con los equipos *time-lapse*, excepto para falsa división, vacuolas y asimetría (utilizando todos los métodos) y multinucleación (para los usuarios de

Primo Vision™), donde esta variabilidad fue más alta. La variabilidad entre laboratorios en catalogación embrionaria según criterios ASEBIR, toma valores de Gwet moderados-sustanciales (Gwet 0.41-0.80) para los usuarios de evaluación morfológica embrionaria clásica (EMEC) y EmbryoScope™, y bajos-moderados (Gwet 0.21-0.60) para usuarios de Primo Vision™. La variabilidad entre laboratorios para decisión clínica fue superior (Gwet 0.41-0.60) en día 5 para usuarios de EMEC que la obtenida para usuarios de equipos *time-lapse* (Gwet 0.81-1).

3.Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad

De manera global e independientemente del producto analizado en la evaluación de fungibles en el programa de CCExt de CT se observó que la S (0.93 y 0.87) para puntas y medios, respectivamente, fue superior a la E (0.79 y 0.76), y que el VPN fue superior (0.92 y 0.90) que el VPP (0.82 y 0.71).

Al comparar los resultados según el material utilizado, observamos de manera global que se obtienen parámetros de rendimiento diagnóstico similares con ambos materiales, excepto en el VPP que resultó significativamente más alto al utilizar puntas de pipeta que medios de cultivo (0,82 [0,77-0,87] vs 0,71 [0,66-0,76], $p \leq 0.05$)

Al analizar las respuestas de los laboratorios por número de participaciones en el programa de CCExt de CT de ASEBIR, se observó que en la primera participación evaluando medios de cultivo, los resultados obtenidos son más bajos en S y E, que los obtenidos en la primera participación evaluando puntas de pipeta. Sin embargo, en las siguientes participaciones, con más experiencia, los resultados en S y E son similares entre ambos productos.

No se observó una correlación significativa entre el porcentaje de aciertos de un laboratorio en medios de cultivo y el de aciertos en puntas de pipeta ($r=0.026$).

Discusión

Acuerdo y fiabilidad entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica

El alto grado de acuerdo observado en este estudio, es similar al observado en este mismo programa, previo a la incorporación de la catalogación estandarizada (Ruiz de Assin et al., 2009). Sin embargo, la tendencia a un mayor desacuerdo entre laboratorios y el panel de expertos en los embriones tipo D, podría deberse a la mayor dificultad en el uso de un sistema de catalogación embrionaria de cuatro categorías.

La fiabilidad “casi perfecta” obtenida en la evaluación de la fragmentación embrionaria, es superior a la publicada hasta el momento (Arce et al., 2006; Paternot et al., 2009, 2011). Esto podría deberse al tipo de imágenes analizadas, ya que en el estudio de Arce et al. (2006) no se utilizaron el mismo tipo de imágenes en los dos grupos de estudio, al contrario de lo que ocurre en este trabajo. Otra causa, podría ser el diferente número de categorías utilizadas en ambos estudios para valorar la fragmentación. En nuestro estudio se utilizan dos categorías para valorar fragmentación mientras que en el estudio de Arce et al. (2006) se utilizan seis categorías para evaluar la misma característica, lo que podría llevar a un menor grado de fiabilidad al analizar fragmentación.

Al analizar presencia/ausencia de multinucleación o de vacuolas en embriones en estadio de células y la catalogación de blastocistos entre laboratorios y un panel de expertos mediante EMEC se obtuvo un porcentaje de acuerdo ponderado casi total entre laboratorios y el panel de expertos, así como una elevada fiabilidad, mientras que al utilizar el coeficiente kappa (k) para medir la fiabilidad, este fue moderado. Esta paradoja de obtener elevados porcentajes de acuerdo con k bajas es debida al análisis de poblaciones muy homogéneas. Esto provoca que el acuerdo obtenido se deba al azar y no ofrezca ninguna garantía de que los embriólogos puedan identificar, estas alteraciones en el embrión.

Los resultados obtenidos de fiabilidad y acuerdo en catalogación de embriones son superiores a los publicados anteriormente (Arce et al., 2006; Baxter Bendus et al., 2006; Paternot et al., 2009; Storr et al., 2016). Esto podría

deberse al diferente diseño de nuestro estudio (en el que la categoría elegida por la mayoría de laboratorios se comparó con la asignada por un panel de expertos) en comparación con el diseño empleado por otros investigadores (variabilidad interobservador).

En la catalogación de blastocisto, se observan altos valores globales de fiabilidad y acuerdo. Sin embargo, observamos al igual que Storr et al., (2017), niveles bajos de fiabilidad entre observadores en la evaluación de la MCI. Nosotros, al tener en cuenta cada categoría, observamos que la baja fiabilidad obtenida es debida principalmente a las MCI de mala calidad.

Con respecto al destino final del embrión, la baja fiabilidad y acuerdo observados, especialmente en la decisión de crioconservar embriones, en comparación a los valores de fiabilidad y acuerdo encontrados en catalogación embrionaria, podría deberse a las diferentes políticas de transferencia y crioconservación propias de cada centro.

La relativa poca fiabilidad observada en el destino del embrión en día 5, en comparación con lo publicado por Storr et al. (2016), podría ser debido a que nuestro programa considera tres opciones de respuesta (transferencia, crioconservar o descartar) mientras que Storr et al. (2016) consideran sólo dos (transferencia o no de transferencia).

Una de las limitaciones de nuestro trabajo es que no disponemos de información sobre el nivel de experiencia de los embriólogos que participan en el CCEExt. Otra posible limitación es que la participación en el programa de CCEExt de ASEBIR es voluntaria.

Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante *time-lapse*

Nuestros resultados, tras analizar los datos de un ensayo piloto de CCEExt de EME mediante *time-lapse*, parecen confirmar la dificultad existente en la determinación de los tiempos a los que ocurren algunos eventos embrionarios debido a la posibilidad de marcar los eventos en distintos momentos (inicial, medio o final), observándose un pico de variabilidad entre laboratorios al analizar eventos que tardan más tiempo en producirse

Otro pico de variabilidad se observa a partir del estadio de 8 células, posiblemente porque las células están más próximas entre sí haciendo más difícil su contaje y porque en estos momentos suele comenzar el proceso de compactación.

El alto grado de acuerdo entre laboratorios, observado al analizar mediante *time-lapse* los eventos de morfocinética coincide con los publicados por Sundvall et al. (2013). En nuestro estudio, la validez de esta conclusión se vio reforzada por la participación de 24 laboratorios, comparado con los tres observadores del mismo laboratorio estudiados por Sundvall et al. (2013).

Tras comprobar los resultados del ensayo piloto de EME mediante *time-lapse* con los obtenidos en el programa de CCExt de EMEC, no se observó un aumento en el nivel de acuerdo entre laboratorios usuarios de equipos *time-lapse* frente a los usuarios de EMEC, en la evaluación de ninguna característica morfológica ni en clasificación de embriones según los criterios ASEBIR. Este resultado inesperado, podría deberse en parte, a que muchas características morfológicas embrionaria son transitorias y los criterios para evaluar dichas características no están bien definidos actualmente para las plataformas *time-lapse* (Hardarson et al., 2006; Chavez et al., 2012).

Entre los usuarios de Primo Vision™ el grado de acuerdo fue significativamente menor en la evaluación de fragmentación y multinucleación posiblemente por limitaciones técnicas de esta plataforma. Por este motivo cualquier nueva tecnología de selección de embriones debe ser evaluada antes de su incorporación a los laboratorios de reproducción asistida (Lundin y Ahlstrom, 2015; Harper et al., 2012).

No se observa una disminución en la variabilidad entre laboratorios en la evaluación de asimetría entre usuarios de equipos *time-lapse* y EMEC. Estos resultados podrían explicarse porque, a pesar de que los equipos *time-lapse* disponen de herramientas para medir el diámetro celular, no todos los laboratorios las utilizan, no todos los laboratorios emplean la misma definición de asimetría y que la asimetría puede ser transitoria (Aparicio-Ruiz et al., 2016).

La variabilidad entre centros en decisión clínica es menor entre usuarios que utilizan equipos *time-lapse*. Este puede deberse a la baja variabilidad

observada en características morfológicas, que sólo pueden ser observados por *time-lapse* (Kirkegaard et al., 2012, Ciray et al., 2014), y que han sido relacionadas con una baja calidad embrionaria (Rubio et al., 2012; Athayde Wirka et al., 2014; Liu et al., 2015).

Algunos autores han sugerido que el uso de algoritmos basados solo en eventos morfocinéticos debe complementarse con la observación de características morfológicas clásicas, ya sea con equipos *time-lapse* (Aparicio-Ruiz et al., 2016) o por EMEC (Kieslinger et al., 2016). Sin embargo, nuestros resultados muestran que esta práctica sigue presentando variabilidad entre laboratorios.

Los resultados de nuestro estudio muestran que los laboratorios de embriología tienden a estar más de acuerdo en la evaluación de embriones de baja calidad que en los de alta calidad y en los embriones que deben descartarse más que en los que deben transferirse, independientemente de la técnica de evaluación embrionaria utilizada, estos resultados coinciden con los publicados por Arce et al. (2006) y Castilla et al. (2010) para los usuarios de EMEC.

Existen distintas limitaciones en este estudio. Los embriones evaluados por cada método son diferentes y han estado sometidos a diferentes condiciones de cultivo. Sin embargo, los tiempos de desarrollo de los embriones analizados fueron similares para cada una de las plataformas.

Otra limitación es que se desconocen las tasas de implantación de los embriones evaluados y, por tanto, no se ha podido determinar, si la variabilidad entre laboratorios, observada en este estudio, en decisión clínica y en catalogación embrionaria puede reducir la efectividad de los ciclos de reproducción asistida. Por último, no se puede descartar la posibilidad del sesgo de participación ya comentada en el estudio anterior.

Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad

En nuestro estudio, se observa una tendencia a obtener resultados de S, E, VPP y VPP analizando puntas de pipeta, superiores a los publicados por Castilla et al. (2010). Esto podría deberse a que, en los años incluidos en nuestro estudio, los centros participantes solo utilizaron el TSE como test y solo analizan

un tipo de material (puntas de pipeta), al contrario de lo que ocurría en el programa de CCExt cuando fue analizado por Castilla et al. (2010), en los que los centros emplearon diferentes métodos de evaluación y se analizaron diversos tipos de materiales tóxicos.

Los resultados en S, E, VPP y VPN obtenidos en nuestro estudio analizando medios de cultivo son inferiores a los observados en el programa de CCExt de CT de la asociación americana de bioanalistas (AAB) (AAB, 2017). Estas diferencias podrían deberse a una diferente manipulación de los medios de cultivo para convertirlos en tóxicos en los dos programas de CCExt de CT analizados.

Los laboratorios participantes detectan mejor los productos citotóxicos que los que no lo son. Desde un punto de vista clínico parece menos perjudicial equivocarse clasificando fungible no tóxico como tóxico que un fungible tóxico como no tóxico. Esta tendencia coincide con la observada por otros trabajos (Castilla et al., 2010; AAB, 2017).

Los mismos laboratorios son más eficaces detectando toxicidad en puntas de pipeta que en medios de cultivo, estas diferencias pueden estar relacionadas con el método utilizado por los laboratorios para evaluar la CT de los fungibles. Cuando se evalúan puntas de pipeta, los espermatozoides control se encuentran en el mismo medio de cultivo que los espermatozoides que están bajo la influencia de la toxicidad de las puntas de pipeta. Sin embargo, cuando se evalúan medios de cultivo los espermatozoides control se encuentran en un medio diferente al que se encuentran los espermatozoides incubados con el medio problema.

Otra posible causa podría ser que las condiciones de los fungibles durante el transporte, almacenamiento y manipulación en los laboratorios hasta su uso, afecte más negativamente a los medios de cultivo que a las puntas de pipeta (Nijs et al., 2009; Swain et al., 2012;).

En nuestro trabajo no hemos ajustado por la experiencia previa de los centros en TSE, pero el hecho de no haber observado relación entre el porcentaje de aciertos en un laboratorio con medios de cultivo y puntas de pipeta

sugiere, que las diferencias observadas se deben más al tipo de producto que a un factor de laboratorio.

Conclusiones

1. La evaluación mediante morfología clásica de embriones sin anomalías morfológicas se realiza de forma fiable en los laboratorios de embriología. Cuando se observan diferencias entre centros y expertos, se deben a la presencia de vacuolas o multinucleación en embriones en estadio de células y a masas celulares internas de baja calidad en estadio de blastocisto.
2. Es necesario mejorar la formación de los embriólogos clínicos en la evaluación de estas características morfológicas al utilizar los métodos clásicos de evaluación embrionaria.
3. Deberían analizarse las políticas de transferencia embrionaria de los centros con el fin de lograr niveles más altos de acuerdo respecto al destino final del embrión cuando es evaluado con métodos morfológicos clásicos, especialmente cuando el destino final de los embriones debe ser crioconservar.
4. La variabilidad entre laboratorios en la determinación de los tiempos a los que ocurren los eventos embrionarios depende de la duración del evento.
5. Los equipos *time-lapse* han disminuido la variabilidad entre laboratorios en la decisión clínica tomada sobre embriones. Sin embargo, no han modificado la variabilidad existente entre laboratorios en la catalogación de embriones.
6. Es necesario estandarizar la interpretación de las alteraciones morfológicas embrionarias observadas en equipos *time-lapse* antes de incorporar estas características a los algoritmos diagnósticos que utilizan dichos equipos.
7. El tipo de fungible empleado en los controles de calidad externos de citotoxicidad del laboratorio de reproducción humana asistida influye en los resultados de estos.
8. La utilización de puntas de pipeta como fungible en un programa de control de calidad externo de citotoxicidad es fácil logísticamente y permite obtener resultados válidos en dichos programas. Sin embargo, no permite

analizar controles de calidad de otras características (pH u osmolaridad) que pueden afectar a la viabilidad de gametos y embriones.

Introducción

1. Técnicas de reproducción asistida

La reproducción asistida engloba un conjunto de técnicas (TRA) encaminadas a facilitar la fecundación del ovocito. Dentro de estas, se encuentran la inseminación artificial (IA), la fecundación in vitro (FIV) y la inyección intracitoplasmática de espermatozoide (ICSI). En la mayoría de las técnicas de reproducción asistida es necesario realizar previamente, tratamientos para la estimulación de la ovulación en la mujer, de tal manera que, en vez de reclutar un solo ovocito por ciclo, se consigue el reclutamiento de varios de ellos, permitiendo aumentar considerablemente, de este modo, la tasa de éxito de las TRA. El seguimiento de la estimulación ovárica se realiza mediante ecografía ovárica vía vaginal (inseminación artificial) o acompañado de determinación de niveles séricos de estradiol (FIV/ICSI).

1.1. Inseminación artificial

En la IA se deposita una alícuota de espermatozoides móviles, en el tracto genital femenino.

Según el tipo de semen utilizado la IA puede denominarse:

- Inseminación artificial conyugal (IAC).
- Inseminación artificial de donante (IAD).

Según el lugar donde se deposite el semen:

- intracervical
- intrauterina
- intratubárica.

En IAC la muestra de semen, suele obtenerse por masturbación el mismo día en que se realiza la inseminación. En la IAD el semen se descongela el mismo día de la inseminación, previa asignación del donante, según las características fenotípicas de la pareja.

La muestra de semen se debe procesar en el laboratorio antes de la inseminación con dos objetivos. Primero, aislar un alto porcentaje de espermatozoides funcionales con una alta movilidad y sin anomalías morfológicas, de otros espermatozoides no viables, células y restos celulares presentes en el semen. Segundo, eliminar el plasma seminal (así como prostaglandinas, agentes infecciosos y otras sustancias presentes en él) que impediría la capacitación espermática.

Entre las técnicas más utilizadas para seleccionar espermatozoides móviles se encuentran los gradientes de densidad y el swim-up.

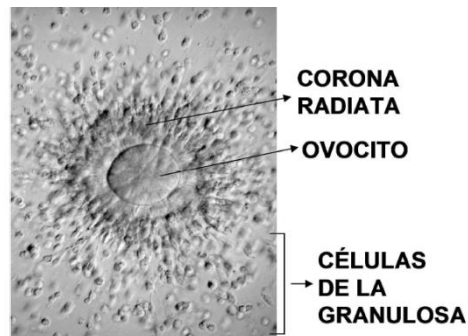
1.2. FIV/ICSI

La fecundación en el laboratorio hace referencia a la fusión de los gametos masculino y femenino y el seguimiento del cigoto generado a lo largo de su desarrollo embrionario, normalmente hasta el día quinto o sexto tras la fecundación. Según el método de fecundación podemos hablar de FIV convencional o de ICSI.

Como se ha comentado anteriormente, para llevar a cabo estas TRA es necesario realizar una estimulación ovárica y posteriormente desencadenar la ovulación. Cuando, tras la estimulación, los folículos ováricos han alcanzado un tamaño de 17-18 mm se procede a desencadenar la ovulación mediante la gonadotropina coriónica humana (hCG) o agonistas de la hormona liberadora de

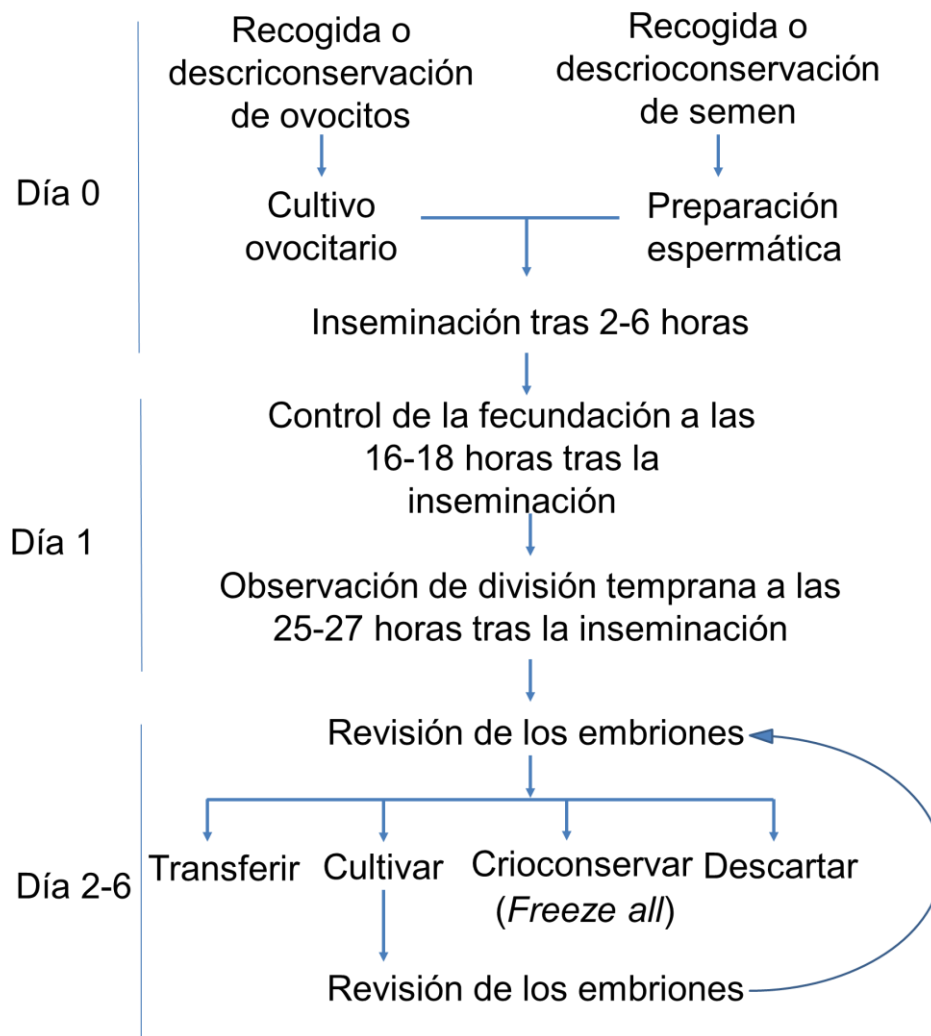
gonadotropinas (GnRH). A las 36 horas de la administración de estas hormonas, se procede a la aspiración de los folículos ováricos, mediante punción folicular ecoguiada. El líquido folicular es trasladado al laboratorio de embriología, donde se realiza la identificación de los complejos cúmulo-corona-ovocito (Figura 1).

Figura 1. Complejo cúmulo-corona-ovocito.



En función de la técnica que se vaya a emplear, FIV o ICSI, el tratamiento del complejo cúmulo-corona-ovocito será diferente. En la FIV convencional se cultiva cada complejo junto a una concentración determinada de espermatozoides, mientras que en la ICSI se decumulan los complejos y cada ovocito será microinyectado con un espermatozoide. A las 16-18 horas, se comprueba la fecundación, realizándose la transferencia de los embriones al útero desde el segundo día hasta el sexto desde que se inseminaron los ovocitos. Los embriones sobrantes se crioconservan en nitrógeno líquido. Otro modo de actuar es crioconservar todos los embriones para una posterior transferencia en diferido, sin realizar transferencia en fresco (*freeze-all*) (Roque, 2015; Ozgur et al., 2017). En la Figura 2 se esquematizan las diferentes etapas de laboratorio en FIV/ICSI.

Figura 2. Resumen de las etapas de laboratorio en FIV/ICSI.



1.3. Otras técnicas

En determinadas situaciones clínicas, se ha sugerido incorporar otras técnicas de laboratorio al protocolo de FIV/ICSI clásico descrito. Entre estas, tenemos la selección avanzada de espermatozoides, la eclosión asistida y el PGT (test genético preimplantacional).

La selección de espermatozoides avanzada se realiza utilizando microscopía óptica de alta amplificación (IMSI) o mediante técnicas biológicas que seleccionan espermatozoides con determinadas características moleculares

(receptores para ácido hialurónico o aquellos que no expresen marcadores apoptóticos).

La eclosión asistida, consiste en la realización de un orificio en la zona pelúcida del embrión que le facilite su abandono en el proceso de eclosión. Dicho orificio puede realizarse mediante láser, de manera mecánica o químicamente (ácido tyrode).

El PGT, consiste en biopsiar una o dos blastómeras de un embrión conseguido tras ICSI en D+3, o un conjunto de células del trofoblasto cuando el embrión está en estadio de blastocisto, para determinar su contenido cromosómico mediante técnicas de hibridación genómica comparativa (CGH) o *next generation sequencing* (NGS) (Bono et al., 2015), o analizar su contenido génico mediante técnicas de biología molecular (reacción en cadena de la polimerasa (PCR), NGS, etc). Dependiendo de la anomalía genética que se pretenda determinar, la técnica se denominará PGT-A (para aneuploidías), PGT-SR (para reordenamientos estructurales cromosómicos) o PGT-M (para defectos monogénicos) (Zegers-Hochschild et al., 2017). Para obtener la biopsia se realiza una apertura en la zona pelúcida del embrión del mismo modo que el indicado anteriormente para la eclosión asistida. También se puede biopsiar el primer o segundo corpúsculo polar, con el objeto de analizar su información genética.

Otra indicación de las técnicas de reproducción asistida es satisfacer los deseos reproductivos de parejas con enfermedades infecciosas transmisibles (VIH, hepatitis C, etc), para reducir al mínimo el riesgo de contagio del otro miembro de la pareja o de la descendencia. En caso del varón seropositivo, a veces, es necesario realizar una doble preparación del semen en el laboratorio

(gradientes de densidad y después swim-up) y analizar la carga viral en esta muestra preparada.

1.4. Vitrificación ovocitaria y embrionaria

Actualmente, la crioconservación embrionaria ha mejorado notablemente sus resultados, debido a que la vitrificación proporciona mejores tasas de supervivencia embrionaria y gestación que la congelación lenta tradicional (Li et al., 2014; Bebrock et al., 2015; Rienzi et al., 2017). Además, los datos que se desprenden de diferentes estudios sobre el huso meiótico de ovocitos descarioconservados con diferentes protocolos de crioconservación sugieren que la vitrificación es el método menos perjudicial para el posterior desarrollo embrionario (Larman et al., 2007; Cobo et al., 2008; Martínez-Burgos et al., 2011).

La principal ventaja de la vitrificación es la ausencia de formación de cristales de hielo. Este método se basa en la solidificación de los solutos a baja temperatura, mediante el aumento extremo de la viscosidad durante el enfriamiento rápido. Los protocolos lentos de congelación se basan en disminuciones graduales de temperatura ($-0,3^{\circ}\text{C}/\text{minuto}$) para que la formación de cristales de hielo sea fuera de la célula. Sin embargo, de esta manera la cantidad de agua del interior celular sólo se reduce de un 70% a un 30%, con lo que aún se pueden formar cristales de hielo en el interior celular.

En cambio, la vitrificación, depende de un descenso rápido de la temperatura $23.000^{\circ}\text{C}/\text{minuto}$ y una alta viscosidad de los crioprotectores, con lo que se evita la formación de cristales en el interior de las células (Al-Hasani et al., 2007; Cobo et al., 2007). La alta osmolaridad del medio de vitrificación

deshidrata rápidamente la célula y su inmersión en el nitrógeno líquido vitrifica la célula antes de que el agua intracelular restante tenga tiempo de formar cristales de hielo perjudiciales. El resultado es que la célula queda en un estado vítreo.

Los estudios de seguimiento de niños nacidos de ovocitos y embriones vitrificados por diversos métodos (Liu et al., 2013; Cobo et al., 2017), han demostrado la seguridad de esta técnica.

2. Evaluación morfológica clásica de ovocitos y embriones

2.1. Evaluación ovocitaria

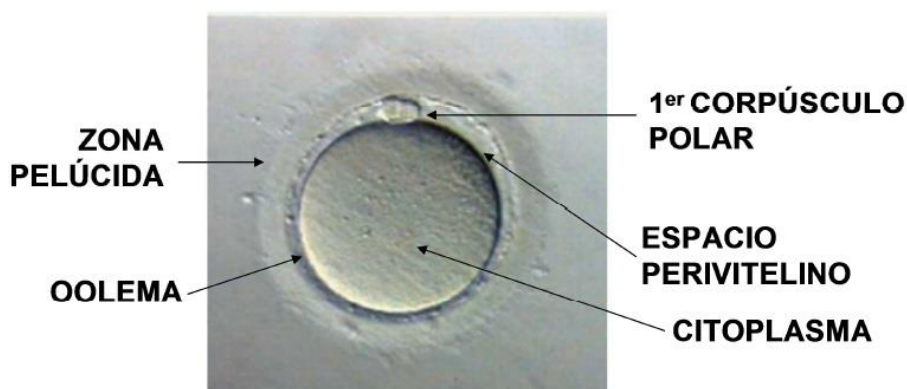
Los ovocitos recuperados de pacientes sometidas a tratamiento de estimulación ovárica para realizar TRA pueden encontrarse en diferentes etapas del desarrollo meiótico. Sólo los ovocitos que se encuentran en el estadio celular de metafase II (morfológicamente se observa que ha extruido el primer corpúsculo polar) son utilizados para ICSI. En cambio, los ovocitos en fase celular metafase I (no se observa el primer corpúsculo polar extruido) o profase I (se observa una vesícula germinal) no pueden ser utilizados para ICSI. Más de la mitad de los ovocitos recuperados en un ciclo de estimulación ovárica muestran una o más características morfológicas alteradas (Ten et al., 2007; Tejera et al., 2011; Magli et al., 2012).

Estas alteraciones morfológicas se pueden subdividir en citoplasmáticas o extracitoplasmáticas. Las primeras incluyen: granulosis localizada en el centro del ovocito, agregación del retículo endoplásmico liso, vacuolización y presencia de inclusiones citoplasmáticas como cuerpos refringentes: cuerpos necróticos o picnóticos. Estas anormalidades reflejarían defectos intrínsecos del ovocito en detrimento de su viabilidad. Las segundas incluyen: Restos celulares en el

espacio perivitelino, anomalías de la zona pelúcida, espacio perivitelino aumentado y alteraciones del primer corpúsculo polar (fragmentación y tamaño). A veces, la morfología del ovocito puede reflejar anomalías genéticas. Por esta razón se aconseja no inseminar los ovocitos gigantes ($>200\mu\text{m}$ de diámetro) ya que suelen contener un set de cromosomas adicional generando fecundaciones anómalas (Hurtado de Mendoza et al., 2015).

Queda aceptada la definición morfológica de un ovocito maduro óptimo, como aquel que ha de tener una forma redondeada, envuelto por una zona pelúcida clara y uniforme con un grosor de alrededor de $20\mu\text{m}$, citoplasma translucido y libre de inclusiones, y con un corpúsculo polar de tamaño apropiado (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011) (Figura 3).

Figura 3. Esquema de un ovocito en metafase II con buena morfología.



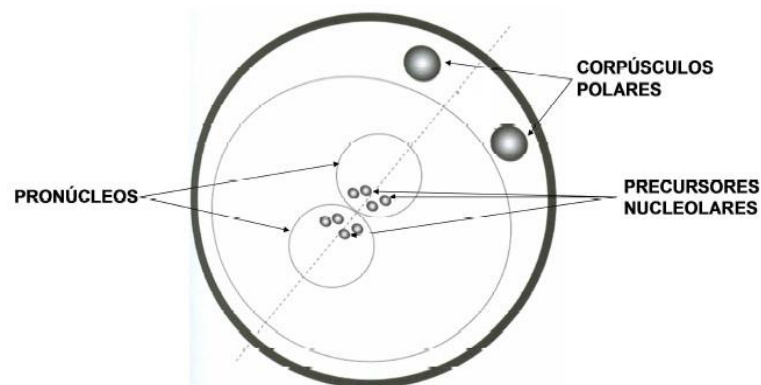
Aunque algunos autores han reportado mayores tasas de embarazo en pacientes con transferencia de embriones procedentes de ovocitos normales morfológicamente frente a pacientes con transferencia de embriones procedentes de ovocitos con alteraciones citoplasmáticas (Alikani y Cohen, 1995; Krisher, 2004; Reader et al., 2017), la revisión bibliográfica realizada revela una gran disparidad en estos resultados (Alikani y Cohen, 1995; De Sutter

et al., 1996; Loutradis et al., 1999; Figueira et al., 2010; Rienzi et al., 2011). El estado actual del conocimiento sobre parámetros morfológicos del ovocito no permite establecer la importancia de dichos parámetros. Los ovocitos obtenidos tras un ciclo de estimulación no pueden mejorarse y todos serán microinyectados con el fin de maximizar la posibilidad de tener embriones útiles. Por esto, la valoración morfológica de los ovocitos no es una práctica muy extendida.

2.2. Evaluación del cigoto

La valoración del cigoto se realiza principalmente mediante la evaluación de parámetros morfológicos de los pronúcleos masculino y femenino, así como de sus precursores nucleolares. De forma generalizada se ha propuesto que esta observación se realice tras 16-18 horas postinseminación (Día 1) (Alpha Scientists In Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011; Hurtado de Mendoza et al., 2015). Sin embargo, algunos autores recomiendan retrasar dicha observación entre 1,5 y 2 horas en los ovocitos inseminados mediante FIV (ESHRE Special Interest Group of Embryology and Alpha Scientists in Reproductive Medicine, 2017) (Figura 4).

FIGURA 4. Esquema de un cigoto.



Los acontecimientos que se han podido observar en el ovocito tras una fecundación normal mediante ICSI se describen a continuación. Tras realizar la ICSI el ovocito muestra “ondas” circulares de granulación en el ooplasma con una periodicidad de 20 a 50 minutos. Durante esta fase de granulación, la cromatina del espermatozoide se descondensa. A continuación, el segundo corpúsculo polar es extruido y el pronúcleo masculino se forma en la zona central del ovocito. Acto seguido se forma el pronúcleo femenino que migra hacia la zona central del ovocito para encontrarse con el pronúcleo masculino. Ambos pronúcleos aumentan su tamaño y los precursores nucleolares se desplazan por el interior de cada pronúcleo para alinearse y quedar enfrentados.

Como hemos comentado, las clasificaciones embrionarias del cigoto publicadas consideran principalmente el aspecto de los pronúcleos (PN) y de los precursores nucleolares (NPB) (Tesarik y Greco, 1999; Tesarik et al., 2000; Scott, 2000). En relación a los PN se valora la simetría (iguales o desiguales), posición (adyacentes o separados), y localización en el citoplasma (centrales o no). Los NPB se clasifican en cuanto a número, simetría (tamaño similar) y localización (situados en la mitad próxima al PN adyacente o polarizados, es decir, cercanos a la membrana puesta al PN adyacente).

Estas clasificaciones fueron evaluadas por otros autores, aunque con pequeñas adaptaciones, relacionándola con calidad embrionaria (Wittemer et al., 2000; Arroyo et al., 2007), ritmo de división y desarrollo hasta blastocisto (Lan et al., 2003; Balaban et al., 2004), implantación (Lan et al., 2003; Balaban et al., 2004) y gestación (Wittemer et al., 2000; Lan et al., 2003; Balaban et al., 2004; Montag et al., 2011). Por el contrario, otras publicaciones no hallan ninguna relación entre esta catalogación y el ritmo de división (James et al. 2006; Aydin

et al., 2011), el desarrollo hasta blastocisto (Guerif et al., 2007) o la tasa de implantación y gestación (James et al. 2006; Arroyo et al., 2007; Brezinova et al., 2009).

Respecto a la morfología del citoplasma se evalúa la presencia de un halo citoplasmático, el cual se produce en parte, por la redistribución de las mitocondrias alrededor de los pronúcleos. Esto origina un anillo claro en la periferia del ooplasma. Diferentes publicaciones concluyen que la presencia de este halo es una característica positiva, siempre y cuando no sea excesiva (Salumets et al., 2001; De Placido et al. 2002; Zollner et al., 2002; Balaban y Urman, 2006).

Como hemos visto, no se ha encontrado consenso en la bibliografía cuando se trata de relacionar los distintos patrones pronucleares con las tasas de implantación, aunque hay tres características morfológicas que sí parecen tener relación con bajas tasas de implantación o con mala morfología embrionaria:

- Un solo precursor nucleolar en un pronúcleo.
- Pronúcleos separados.
- Pronúcleos de tamaño desigual.

La presencia de estas características se ha relacionado además con mayores tasas de aneuploidías y alteraciones en la dotación cromosómica del embrión (Hurtado de Mendoza et al., 2015).

2.3. Evaluación embrionaria

La valoración morfológica del preembrión ha constituido tradicionalmente la base de la determinación de la calidad embrionaria. Para realizar esta valoración el embrión se observa la presencia de división temprana (DT) entre las 25-27 post ICSI (Día 1), también se observa en el intervalo de 43-45 horas post inseminación (Día 2) y 67-69 post inseminación (Día 3) (Hurtado de Mendoza et al., 2015).

Se puede realizar una segunda evaluación embrionaria en día 1 del desarrollo, a las 25-27 horas post-inseminación para evaluar la presencia o ausencia de los pronúcleos y/o la división temprana a dos células (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011).

La relación entre los embriones con DT y la calidad embrionaria (Lundin et al., 2001; Rienzi et al., 2005; Ciray et al., 2006), el desarrollo hasta blastocisto (Fenwick et al., 2002; Neuber et al., 2003; Guerif et al., 2007), la viabilidad embrionaria (Shoukir et al., 1997; Salumets et al., 2003) y la tasa de implantación (Rienzi et al., 2005; Ciray et al., 2006) son controvertidas. Además, en el caso de las tasas de gestación, algunos autores han relacionado la DT con un aumento de esta tasa, tanto en FIV (Shoukir et al., 1997) como en ICSI (Ciray et al., 2005), en otras publicaciones no se confirma esta relación (Emiliani et al., 2006; Sundström y Saldeen, 2008; De los Santos et al., 2014).

Por tanto, el uso de este parámetro parece no relacionarse con la tasa de implantación ni la tasa de recién nacidos, por lo que se usa como un criterio

adicional únicamente cuando la calidad embrionaria es subóptima (De los Santos et al., 2014).

En morfología embrionaria clásica, los criterios más utilizados para catalogar los embriones en día 2 y 3 del desarrollo, según su potencial de implantación son: número de células y ritmo de división, simetría celular y sincronía de división celular, porcentaje y tipo de fragmentación celular, visualización de núcleos y grado de multinucleación y presencia de vacuolas. En menor medida también se usan como criterios para valorar el potencial de implantación: presencia de anillo acitoplasmático, *pitting* o moteado, formas de las blastómeras, morfología de la zona pelúcida y el grado de compactación/adhesión temprana.

Número de células y ritmo de división

La cinética de división embrionaria hace referencia al número de células que se pueden observar en el embrión en sus diferentes estadios de desarrollo. Estos valores esperados quedan reflejados en la Tabla 1.

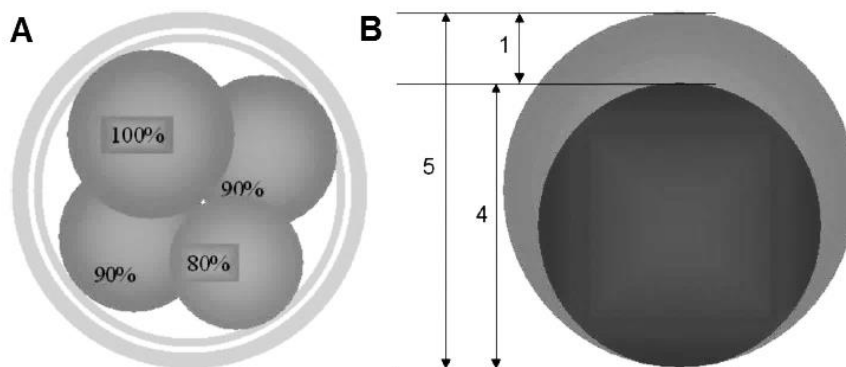
Tabla 1. Cinética del desarrollo embrionario óptimo post-inseminación esperada en FIV/ICSI usando evaluación clásica (Hurtado de Mendoza et al., 2015).

Cigoto	16-18 horas	2 pronúcleos
División temprana	25-27 horas	2 células estadio específico
Día 2	43-45 horas	4 células estadio específico
Día 3	67-69 horas	8 o 7 células estadio específico
Día 4	90-94 horas	Mórula
Día 5	114-118 horas	Blastocisto temprano
Día 6	136-140 horas	Blastocisto expandido

Simetría de las blastómeras

El tamaño de las blastómeras también es un factor útil para estudiar el potencial de desarrollo e implantación de los embriones (Hardarson et al., 2001). Una división desigual origina blastómeras de diferente tamaño con distinta distribución del material genético, lo cual repercute negativamente en el potencial de implantación (Steer et al., 1992; Van Blerkom et al., 2000; De Placido et al., 2002; Hnida et al., 2004). Según Hardarson et al. (2001), un preembrión de 4 células con división asimétrica es aquel en el que la diferencia entre el diámetro de las blastómeras mayor y menor supera el 20%. En la figura 5 se muestra un esquema de un embrión que presenta asimetría entre sus blastómeras.

Figura 5. Asimetría en la división celular embrionaria. La figura 5.A muestra un embrión en el que existe una diferencia de volumen del 20% entre sus blastómeras. En la figura 5.B se superponen 2 blastómeras con una diferencia de tamaño del 20%. (Hurtado de Mendoza et al., 2015).



No obstante, a veces la asimetría entre blastómeras puede ser normal, pues se debe a una asincronía en la división celular. En la Figura 6 se muestra la asimetría embrionaria que corresponde con asincronía en la división celular para embriones en día 2 y día 3 (Holte et al., 2007).

Otro aspecto de la división celular que se puede evaluar en el embrión en día 2 es el plano de división de las blastómeras. El embrión presentará una determinada disposición de las blastómeras en función de la orientación de los planos de división para pasar de 2 a 4 células (Figura 7).

Figura 6. Embriones con asimetría entre sus blastómeras por asincronía en la división celular.

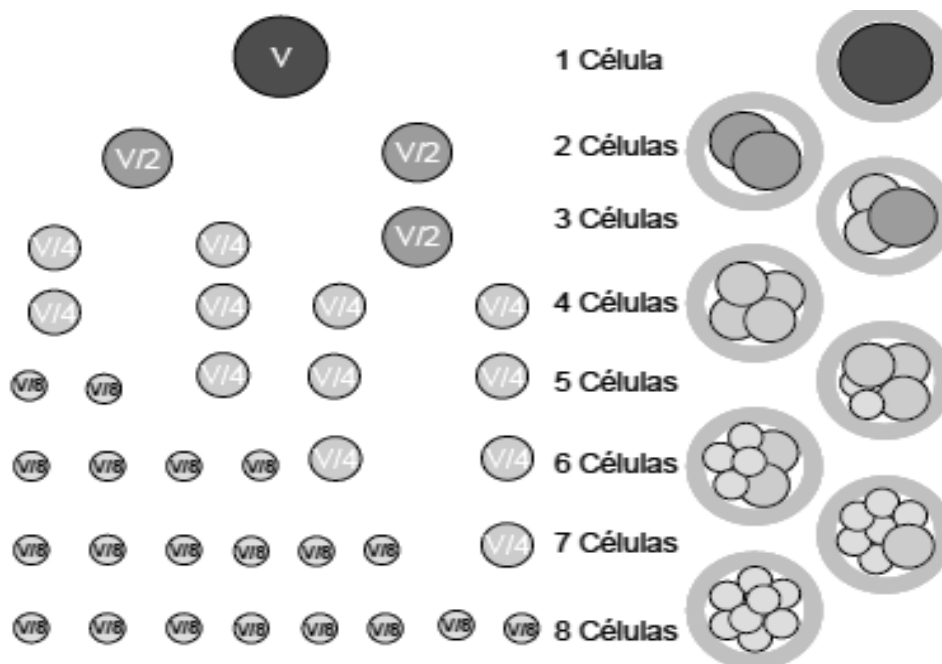
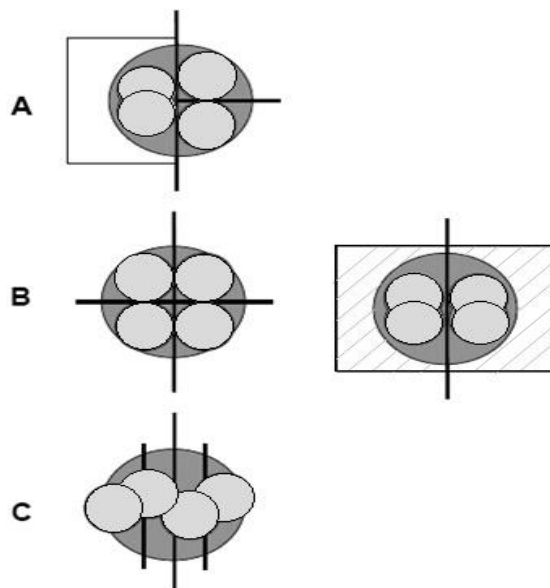


Figura 7. Planos de división embrionaria.



Holte et al. (2007) encontraron mayores tasas de implantación cuando los planos de la segunda división son perpendiculares entre sí, como se muestra en la figura 7.A. En la figura 7.B y 7.C los planos de la segunda división celular son paralelos entre sí, y perpendiculares (7.B) o paralelos (7.C) al plano de la primera división celular.

Porcentaje y tipo de fragmentación celular

Es necesario evaluar el tamaño, distribución y porcentaje de fragmentación citoplasmática.

Lo primero, es saber distinguir una blastómera de un fragmento citoplasmático. Johansson et al. (2003) determinó el diámetro normal de las blastómeras, resultando de 65-70 μm en día 2 y algo menor en día 3 (55-60 μm). Además del tamaño, la característica fundamental para identificar una blastómera será la presencia de uno o varios núcleos. Sin embargo, aunque un fragmento es una estructura formada por citoplasma delimitado por una membrana, es una estructura anuclear y con un tamaño inferior a 45 μm en día 2 y 40 μm en día 3 (Prados et al., 2012).

En lo que respecta al tamaño y distribución de los fragmentos, son los de gran tamaño y los repartidos por todo el embrión los que mayor influencia negativa tienen sobre la tasa de implantación (Alikani y Cohen, 1995; Alikani et al., 2000). Respecto al porcentaje de fragmentación, no repercute en la tasa de implantación si es inferior al 20-25% (Ziebe et al., 1997; Van Royen et al., 1999; Alikani et al., 2000; Hardarson et al., 2001; Racowsky et al., 2003).

Visualización de núcleos y grado de multinucleación

La multinucleación se define como la presencia de más de un núcleo en al menos una de las blastómeras de un embrión (Hurtado de Mendoza et al., 2015).

Mientras que la visualización de un núcleo único por blastómera es un buen predictor del potencial de implantación, la presencia de blastómeras multinucleadas implica bajo potencial de implantación y aumento en la tasa de aborto (Meriano et al., 2004; Saldeen y Sundström, 2005; Cuadros et al., 2011). Además, la multinucleación se correlaciona de forma directa con el incremento en la tasa de anomalías cromosómicas embrionarias (Hardarson et al., 2001).

A pesar de la importancia que tiene la observación del número de núcleos por blastómera, esta observación es difícil debido a que los núcleos no están presentes durante todo el desarrollo embrionario, si no que aparecen y desaparecen en función de la fase mitótica de la blastómera. El desarrollo del embrión también dificulta la observación de núcleos, ya que cada vez hay un mayor número de blastómeras y estas van siendo más pequeñas. Estas dificultades, generan variaciones significativas de la valoración del grado de nucleación entre observadores como comentaremos posteriormente (Arce et al., 2006; Paternot et al., 2009).

Aunque se han publicado nacimientos de niños sanos (Fauque et al., 2013), las sociedades científicas recomiendan no seleccionar los embriones multinucleados para transferir, si existen embriones no multinucleados (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011).

Presencia de vacuolas

Las vacuolas se definen como inclusiones citoplasmáticas rodeadas de membrana y rellenas de un fluido virtualmente idéntico al del espacio perivitelino. La presencia algunas vacuolas de tamaño pequeño (5-14 μm) no compromete el desarrollo embrionario, al contrario de lo que ocurre cuando la vacuolización es muy extensa (Prados et al., 2012).

Sin embargo, como ocurría con la multinucleación, es difícil cuantificar el grado de afectación por vacuolas que presenta un embrión, ya que no solo varía el número sino también el tamaño de éstas a lo largo del desarrollo embrionario (Hurtado de Mendoza et al., 2015).

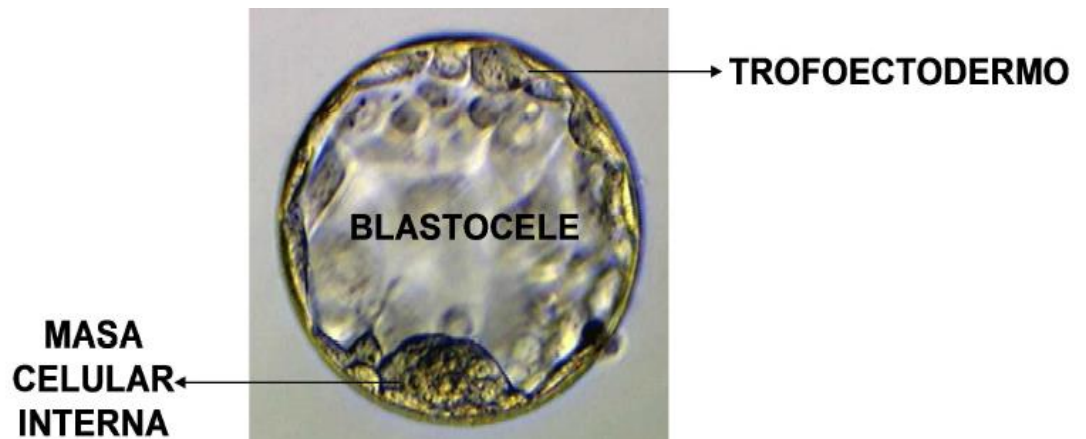
2.4. Evaluación de blastocistos

El cultivo de embriones hasta estadio de blastocisto es una técnica no invasiva de selección embrionaria que puede contribuir a una mejor selección.

El desarrollo embrionario se encuentra sujeto a los transcritos de origen materno hasta día 3, momento a partir del cual se activa el genoma embrionario. Únicamente cultivando los embriones hasta estadio de blastocisto podremos identificar aquellos embriones con bloqueo de desarrollo en día 3.

El estado del embrión en día 5 ó 6 de desarrollo puede variar desde una etapa más retardada de compactación hasta una etapa más avanzada de blastocisto expandido. La calidad de este último se determina por la formación de una cavidad diferenciada rellena de líquido (blastocelo), la capa de células externas que rodean a esta cavidad (trofoectodermo) y el conjunto de células internas que conforman la masa celular interna (MCI) (Figura 8).

Figura 8. Esquema de un blastocisto expandido.



Actualmente, existe cierta controversia sobre qué variable es la mejor para definir el potencial de implantación, si el grado de expansión del blastocelo (Lagalla et al., 2015; Richardson et al., 2015), el tamaño de la masa celular interna (Richter et al., 2001) o las características del trofoectodermo (Ahlstrom et al., 2011; Hill et al., 2013). Aunque los tres parámetros han demostrado su relación con el potencial de implantación y la tasa de nacido vivo (Van den Abbeel et al., 2013), sólo el trofoectodermo parece ser un predictor estadísticamente significativo de nacidos vivos tras ajustar por distintos factores de confusión. Aunque la MCI es importante, un trofoectodermo cohesionado es esencial en esta etapa embrionaria, permitiendo la eclosión exitosa y la implantación (Ahlstrom et al., 2011).

2.5. Sistemas de catalogación embrionaria

Numerosos sistemas de catalogación embrionaria han sido creados teniendo en cuenta diversos aspectos de los comentados anteriormente, con la finalidad de clasificar los embriones en función de su potencial de implantación (Cummins et al., 1986; Hill et al., 1989; Steer et al., 1992; Giorgetti et al., 1995; Bras et al., 1996; Rijnders et al., 1998; Avery y Brinsden, 1999; Alpha Scientists

in Reproductive Medicine and ESHRE Special Interest Group of Embryology., 2011; Montag et al., 2011). Sin embargo, aunque existe cierto acuerdo sobre qué es un buen embrión y qué es un mal embrión, la ausencia de criterios comunes afecta, tanto a los parámetros a evaluar, como a los puntos de corte entre categorías.

Con la intención de unificar criterios de evaluación embrionaria la Asociación para el Estudio de la Biología de la Reproducción (ASEBIR) desarrolló un sistema de gradación de la calidad embrionaria en día 2, 3, 4, 5 y 6 de desarrollo embrionario (Ardoy et al., 2008; Hurtado de Mendoza et al., 2015). En las figuras 9 y 10 se muestran los esquemas de esta catalogación morfológica para embriones en día 2 y día 3 (basándose principalmente en número de células, simetría, porcentaje de fragmentación, multinucleación y presencia de vacuolas). Así, los embriones se clasifican en 4 categorías (A, B, C, D) que se relaciona directamente con su potencial de implantación. En la figura 11 se muestra el esquema para embriones en día 4 y en las figuras 12 y 13 se muestran los esquemas para embriones en día 5 y día 6, basados en la calidad de la MCI y el trofoectodermo.

Figura 9. Catalogación embrionaria en día 2 (Hurtado de Mendoza et al., 2015).

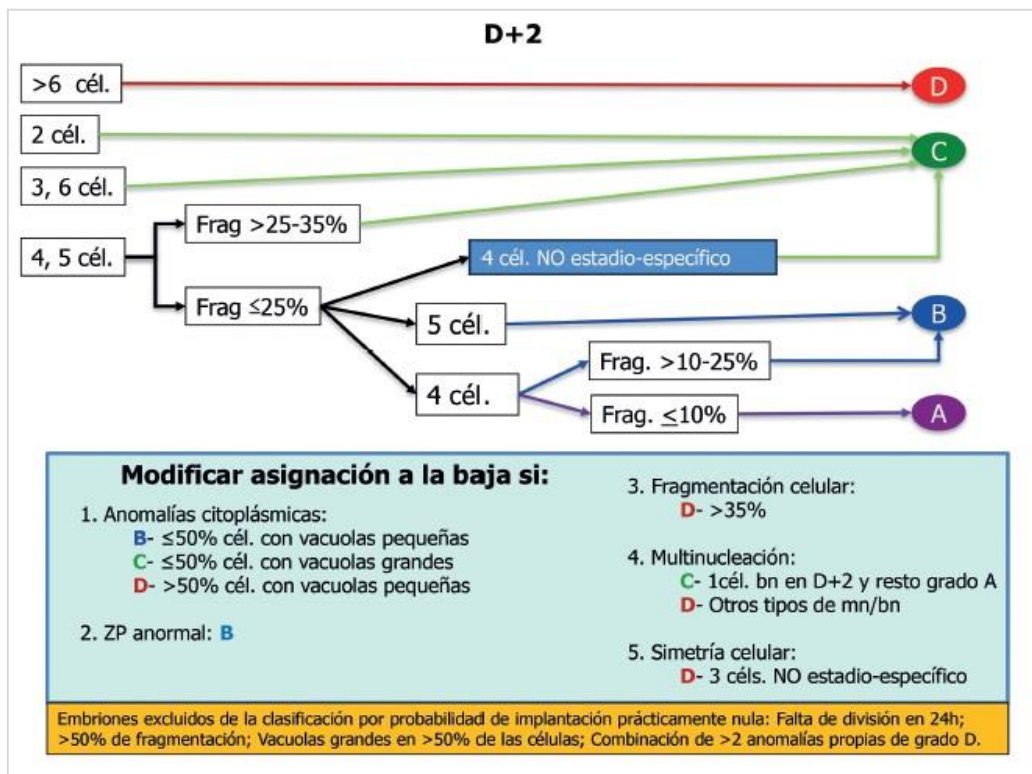


Figura 10. Catalogación embrionaria en día 3 (Hurtado de Mendoza et al., 2015).

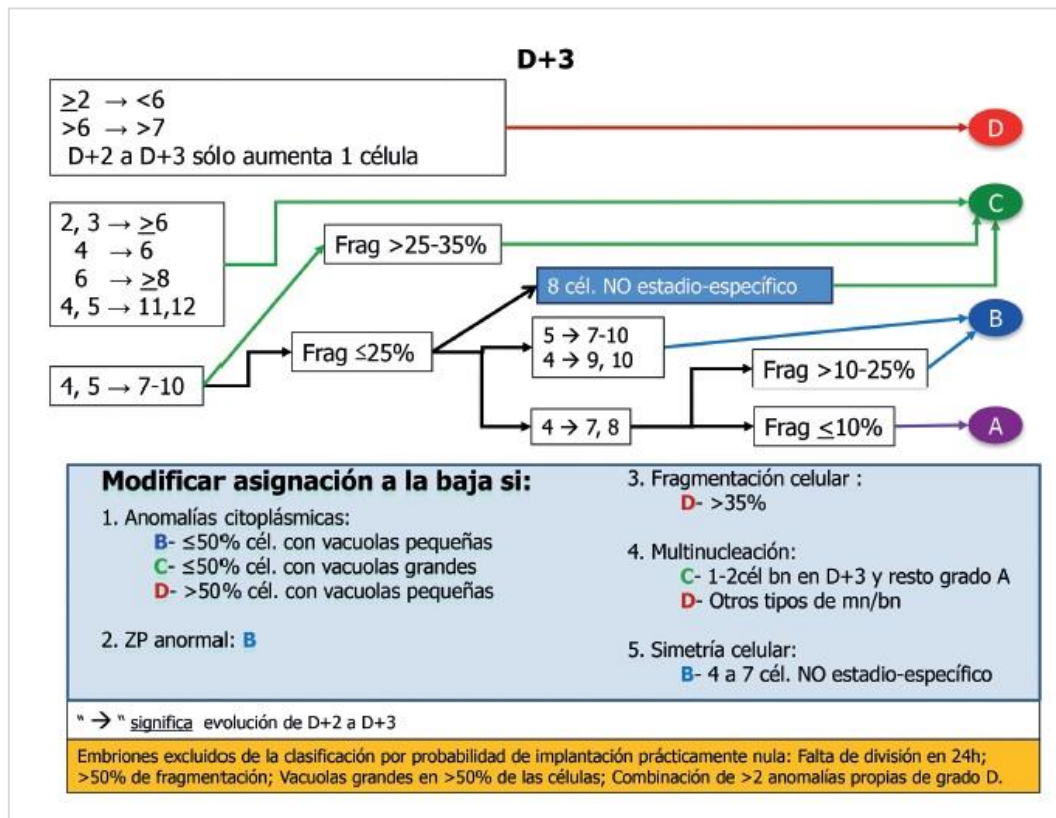


Figura 11. Catalogación embrionaria en día 4 (Hurtado de Mendoza et al., 2015).

D+3	Características morfológicas D+4	D+4
A	Cavitación temprana Compactación total y >8 células	A
	Compactación parcial (1-2 cél. excluidas)	B
	Compactación parcial (>2 cél. excluidas) No compactación (> 8 cél.)	C
B	Cavitación temprana Compactación total y >8 células Compactación parcial (1-2 cél. excluidas)	B
	Compactación parcial (>2 cél. excluidas) No compactación (> 8 cél.)	C
C	Cavitación temprana Compactación total y \geq 8 células	C
D	Compactación parcial	D
	Cualquier característica	

Cualquier embrión que presente en D+4: <ul style="list-style-type: none"> Fragmentación celular >35% Excesiva vacuolización. \leq8 células sin signos de compactación o con compactación de <50% del embrión. 	→	D
--	---	---

Embriones excluidos de la clasificación por probabilidad de implantación prácticamente nula: Falta de división en 24 horas y embriones que presentan una combinación de >2 características propias de la categoría D.

Figura 12. Catalogación embrionaria en día 5 (Hurtado de Mendoza et al., 2015).

D+4	D+5			
	Grado de expansión	MCI	Trofoectodermo	ASEBIR
Mórula compacta	Desde: "Iniciando la expansión" Hasta: "Eclosionando"	A	A	A
			B	B
			C	C
			D	D
		B	A	A
			B	B
			C	C
			D	D
		C	A	A
			B	B
			C	C
			D	D
		D	A,B,C o D	D
Blastocisto temprano o cavitando (ZP gruesa)			C	
Mórula no compacta	Mórula	D		

Figura 13. Sistema de catalogación embrionaria en día 6 (Hurtado de Mendoza et al., 2015).

D+5	D+6			
	Grado de expansión	MCI	Trofoectodermo	ASEBIR
Blastocisto temprano o cavitando (ZP gruesa)	Desde: "Iniciando la expansión"	A	A	B
			B	B
			C	C
			D	D
		B	A	B
			B	B
	Hasta: "Eclosionando"	C	C	C
			D	D
			A	B
			B	B
		D	C	C
			D	D
Blastocisto temprano o cavitando (ZP gruesa)			D	
Mórula no compacta	Mórula		Excluidos de la Clasificación	

La evaluación morfológica clásica es un sistema ampliamente implantado en el laboratorio de reproducción asistida, estando los embriólogos clínicos familiarizados con su uso. Además, existen programas de control de calidad externo (CCExt) desde hace varios años que aumentan su validez (Ruíz-Assin et al., 2009; Castilla et al., 2010). Sin embargo, de la evaluación morfológica basada en observaciones estáticas merece la pena destacar tres grandes desventajas:

En primer lugar, la microscopía estática está vinculada a momentos puntuales, observaciones que deben durar menos de un minuto, con la consiguiente pérdida de información sobre el resto del desarrollo embrionario.

En segundo lugar, la microscopía estática inevitablemente perturba las condiciones óptimas de cultivo, ya que es necesario sacar el embrión de su

incubadora en la que se encuentra en condiciones favorables para su desarrollo, para someterlo a condiciones ambientales del laboratorio mientras se observa en el microscopio y por último, la microscopía estática solo permite interpretaciones cualitativas altamente subjetivas y propensas a variaciones intra e interobservador (Arce et al., 2006; Castilla et al., 2010).

3. Nuevas técnicas de evaluación embrionaria: Equipos *time-lapse*

Se trata de una técnica no invasiva que minimiza la manipulación de los embriones durante su cultivo, y se basa en el uso de la tecnología *time-lapse*. La captura de imágenes se realiza en múltiples planos focales a intervalos de tiempo programables. La suma de fotogramas de cada embrión genera una grabación que permite visualizar el desarrollo embrionario de manera continua, en forma de video. De este modo se registra cada uno de los eventos que suceden durante el desarrollo embrionario aportando certeza y objetividad al criterio, además de poder determinar con precisión el momento exacto en que éstos han ocurrido (Lemmen et al., 2008). De esta forma, la catalogación morfológica clásica se aúna con la cinética de división celular, en una nueva forma de catalogación embrionaria denominada, catalogación morfocinética.

3.1. Equipos *time-lapse*

Existen diferentes plataformas de microscopía *time-lapse*, entre las que podemos encontrar:

- Primo Vision™ (Vitrolife): Microscopio que se puede incorporar a un incubador convencional. Presenta una cámara diseñada para obtener imágenes automáticas en los tiempos previamente definidos a lo largo de 11 planos

focales. Tiene capacidad de albergar un máximo de 16 embriones al mismo tiempo.

- Eeva™ (Merck): Mediante iluminación de campo oscuro captura imágenes en un único plano focal. Analiza de forma automática los tiempos de división con la intención de determinar el potencial de desarrollo del embrión en día 3 y seleccionar directamente los embriones de mejor pronóstico, basándose en los trabajos de Wong (Wong et al., 2010).

- Embryoscope+™ (Vitrolife): Incubador con sistema *time-lapse* integrado, toma imágenes a lo largo de 7 planos focales. Tiene capacidad para 240 embriones al mismo tiempo, y utiliza un software de análisis de imagen en el que todos los eventos embrionarios considerados se pueden interpretar fácilmente.

- Geri® (Merck): Es un incubador que tiene incorporadas seis cámaras individuales de alta resolución con enfoque modular en 11 planos focales, en compartimentos con apertura *benchtop*. Puede albergar hasta 96 embriones. Incorpora distintos algoritmos morfocinéticos.

- Miri® (Esco Medical): Se trata de un incubador con seis cámaras incorporadas. Los pequeños volúmenes de la cámara y la regulación de calor directo se traducen en una recuperación rápida de la temperatura y del gas.

3.2. Ventajas y limitaciones de los equipos *time-lapse*

Con la morfocinética, la cantidad de información que se obtiene es mucho mayor y teóricamente más objetiva, ya que se pueden cuantificar todos los eventos celulares como la aparición y desaparición de los PN, tiempo exacto de

cada división celular y duración de las diferentes citocinesis (Wong et al., 2010; Meseguer et al., 2011; Aparicio et al., 2013; Aguilar et al., 2014).

Entre las ventajas de los equipos *time-lapse* podemos encontrar:

Esta nueva tecnología permite el mantenimiento de condiciones óptimas de cultivo durante todo el desarrollo embrionario ya que el pH, la temperatura y la humedad no se ven alterados. Esto es imposible de conseguir mientras se realizan las observaciones de los embriones mediante microscopía convencional. Además, se realiza una monitorización continua de los embriones, ya que los mismos son cultivados directamente bajo una cámara. Esto nos permite realizar determinaciones precisas del momento en el que ocurren cada una de las divisiones celulares y una observación más cercana de acontecimientos morfológicos tales como la aparición y desaparición de pronúcleos o el inicio de la primera división.

Otra ventaja que podemos mencionar es la reducción de posibles errores en la catalogación de embriones. El aspecto morfológico de un embrión puede cambiar en un periodo muy corto de tiempo lo cual conduce a decisiones diferentes con respecto al tipo y número de embriones a transferir. También cabe destacar una mejor gestión del tiempo, ya que no es necesario estar presente en el laboratorio para realizar las evaluaciones embrionarias puntuales que requiere la valoración morfológica estática y una mejor gestión de los recursos. La mejora en las condiciones de cultivo y la disponibilidad de información más objetiva, precisa, cualitativa y cuantitativa representan las principales ventajas de estos sistemas.

No obstante, los equipos *time-lapse* también presentan algunas limitaciones. Al ser una tecnología de reciente implantación, aun encontramos inconsistencias en la terminología utilizada a la hora de nombrar las variables morfocinéticas. Kaser y Racowsky (2014) pusieron de relieve que el mismo evento del desarrollo embrionario era denominado de hasta cuatro maneras diferentes según los autores. Esta terminología debería ser estandarizada (Ciray et al., 2014; Kaser y Racowsky, 2014).

Además, se pueden definir unas limitaciones técnicas y metodológicas.

Limitaciones técnicas: Durante el cultivo se pueden generar burbujas de aire, debido a las características de los medios de cultivo, que pueden hacer que la imagen se pierda durante algún tiempo si se sitúan encima del embrión.

Limitaciones metodológicas: En cuanto a los tiempos de división, no son absolutos, sino que oscilan en un rango. Esto es debido a que el tiempo de la microinyección, denominado tiempo cero (t_0), no es el momento exacto en el que se microinyecta cada ovocito, sino la media de tiempo en el que se ha realizado la ICSI. Según algunos autores, estos desfases son muy pequeños y no afectaría a los resultados obtenidos (Ciray et al., 2014). Por el contrario, otros autores consideran que estas diferencias si son significativas. Liu et al. (2015) observaron un retraso en las divisiones de embriones procedentes de FIV al compararlos con embriones procedentes de ICSI, lo que afectaría a una incorrecta determinación del tiempo al que ocurre cada evento, provocando, por tanto, una mala catalogación embrionaria. Esta diferencia en los tiempos de división entre embriones procedentes de FIV o ICSI desaparece si tomamos como tiempo cero la desaparición de los pronúcleos (Liu et al., 2015).

Además, existen muchos factores (medio de cultivo, estimulación ovárica, etiología de la esterilidad, edad materna, etc.) que pueden hacer variar la cinética embrionaria, lo que dificulta su interpretación (Ciray et al., 2012; Kirkegaard et al., 2013). Existen estudios que indican que las variables cinéticas que más se influyen por el medio de cultivo empleado son los tiempos de división celular, y no los intervalos de tiempo que duran estas divisiones celulares (Basile et al., 2013).

3.3. Parámetros cinéticos embrionarios

Una de las estrategias más desarrolladas en el establecimiento de nuevos marcadores de selección, es la observación de los tiempos de división embrionaria. Dado su facilidad de aplicación y su mayor objetividad, gran parte de la investigación reciente se centra en explorar la utilidad de parámetros morfocinéticos como marcadores de competencia embrionaria (Meseguer et al., 2011; Basile et al., 2015).

Los parámetros cinéticos harían que la selección embrionaria sea mucho más fiable (Dal Canto et al., 2012), ya que, si comparamos dos embriones de 4 células de un mismo ciclo, pueden parecer similares en una observación puntual, pero en una evaluación dinámica pueden ser muy diferentes, dependiendo de cómo y cuándo han sido sus divisiones celulares. Actualmente, se han realizado varios estudios y revisiones (Meseguer et al., 2011; Rubio et al., 2012; Hlinka et al., 2012; Azzarello et al., 2012; Chen et al., 2013; Campbell et al., 2013; Kaser y Racowsky, 2014) sobre qué parámetros morfocinéticos se relacionan con el desarrollo hasta blastocisto, la tasa de implantación o gestación clínica.

En la Figura 15 se muestra un resumen de los eventos cinéticos analizados por Kaser y Racowsky (2014). Se trata de un esquema del desarrollo embrionario con los intervalos de tiempo de las variables morfocinéticas predictoras de la llegada a blastocisto e implantación en 9 de los 13 estudios que evaluaron durante su metaanálisis. Estos trabajos, han presentado resultados dispares pues los parámetros morfocinéticos incluidos en los algoritmos diagnósticos son muy diferentes entre los distintos autores, ya sea en predicción de formación de blastocistos (Wong et al., 2010; Cruz et al., 2012; Kirkegaard et al., 2013; Milewski et al., 2015) como en tasa de implantación (Dal Canto et al., 2012; Hashimoto et al., 2012; Chamayou et al., 2013; Desai et al., 2014). Además, otros autores como Kirkegaard et al. (2013) o Desai et al. (2014) no encuentran relación entre morfocinética embrionaria y resultados reproductivos.

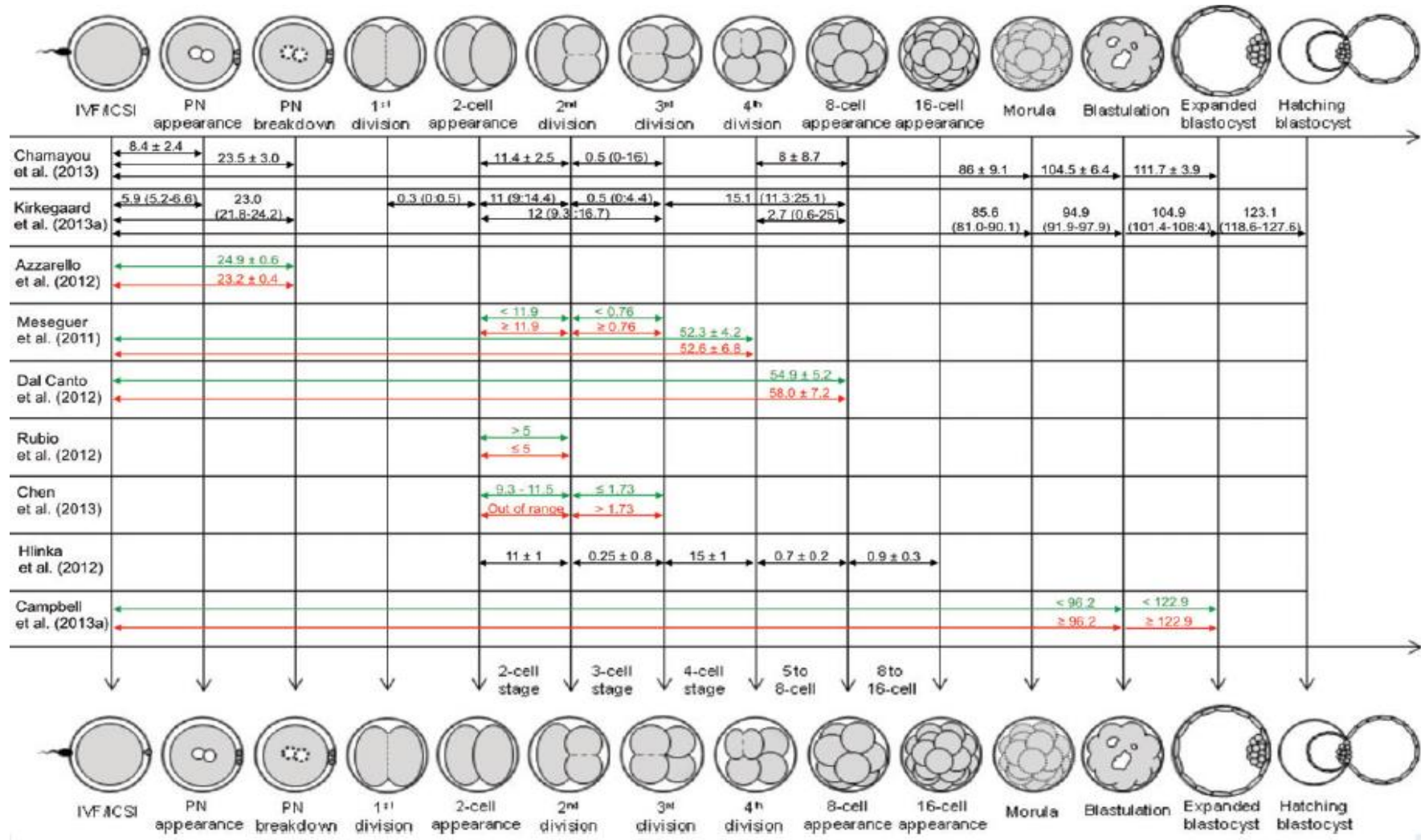
Los parámetros cinéticos identificados por *time-lapse* se dividen en parámetros morfocinéticos directos e indirectos. Definiendo los parámetros morfocinéticos directos como aquellos tiempos en los que ocurre un evento del desarrollo embrionario (extrusión del 2º corpúsculo polar, aparición de PN, desaparición de PN, divisiones celulares...) (Tabla 2) y los indirectos, como aquellos intervalos de tiempo que pueden calcularse relacionando los parámetros morfocinéticos directos (Tabla 3).

a) Parámetros morfocinéticos directos:

Payne et al., (1997) informaron que los embriones de buena calidad parecían proceder de ovocitos con un tiempo similar desde la ICSI hasta la desaparición de PN y que, además, tendían a manifestar ondas en el citoplasma más prolongadas. Más tarde, diferentes estudios evaluaron la aparición, singamia y desaparición de los PN (Chamayou et al., 2013; Kirkegaard et al.,

2013) pero no encontraron diferencias entre los embriones que implantan y los que no. Sin embargo, Azzarello et al., (2012) mostraron que el tiempo de desaparición de los PN es significativamente mayor en aquellos embriones en los que se obtiene un recién nacido (24,9 vs 23,3 h) y que la desaparición de los mismos antes de 20,5 h implica no implantación o aborto (Azzarello et al., 2012). Otros autores han encontrado que la capacidad de implantación disminuye en aquellos cigotos en los que los PN desaparecen antes de 22,2 h o después de 25,9 h tras la ICSI, así como cuando ocurre una rápida extrusión del 2º corpúsculo polar (CP) (1-3,2 h) (Aguilar et al., 2014).

Figura 15. Parámetros cinéticos identificados por *time-lapse* como predictores de desarrollo embrionario e implantación según Kaser y Racowsky (2014).



Gracias a la tecnología *time-lapse* se ha podido correlacionar la división temprana de forma positiva, definiéndose ésta como la primera división mitótica ocurrida entre las 25-27 h post microinyección. La DT se ha relacionado con un mayor número de células en día 2 (4 células o más), así como con mejores tasas de implantación (Lemmen et al., 2008). Sin embargo, si la DT ocurre de manera precipitada, antes de 24,3 h, se ha relacionado con menores tasas de implantación (Meseguer et al., 2011).

Algunos estudios han analizado los tiempos de división tempranos, encontrando que un desarrollo rápido en las primeras divisiones celulares favorece el desarrollo óptimo hasta día 5 (Hashimoto et al, 2012; Chamayou et al., 2013; Herrero et al., 2013; Desai et al., 2014). Por otro lado, las divisiones lentas comprometen el desarrollo hasta día 3. Los autores concluyen que parámetros tempranos (t_2 , t_3 , t_4) están relacionados con el desarrollo hasta día 3, y se necesitan además t_5 y t_8 para predecir la formación de blastocisto (Herrero et al., 2013). Otros autores concluyen que embriones que se bloquean en estadios avanzados se dividen hasta 6 células con una cinética similar a embriones que se desarrollan a blastocisto, pero se produce un retraso significativo en t_8 (Dal Canto et al., 2012).

Tabla 2. Parámetros morfocinéticos directos (Ciray et al., 2014).

Tiempo	Definición de eventos
tPB2	Extrusión del 2º corpúsculo polar.
tPNa	Aparición de pronúcleos.
tPNf	Desaparición de pronúcleos.
tSt2	Inicio de t2.
t2	Tiempo en el que embrión presenta 2 células.
t3	Tiempo en el que embrión presenta 3 células.
t4	Tiempo en el que embrión presenta 4 células.
t5	Tiempo en el que embrión presenta 5 células.
t6	Tiempo en el que embrión presenta 6 células.
t7	Tiempo en el que embrión presenta 7 células.
t8	Tiempo en el que embrión presenta 8 células.
t9	Tiempo en el que embrión presenta 9 células.
tM	Tiempo en el que alcanza la fase de mórula
tSB	Tiempo en el que aparecen los primeros signos de blastulación
tB	Tiempo en el que alcanza la fase de blastocisto
tE	Tiempo en el que comienza la expansión del blastocisto
tHN	Tiempo en el que inicia la eclosión del blastocisto
tH	Tiempo en el que ha eclosionado del blastocisto

b) Parámetros morfocinéticos indirectos:

Por otro lado, con los equipos *time-lapse* se abre un nuevo método de evaluación de los embriones, ya que no sólo se puede conocer en qué momento ocurren las distintas divisiones celulares, sino que también, puede calcularse la duración de cada ciclo celular. Especialmente, el intervalo de tiempo de 2 a 3 células, 3 a 4 células, 4 a 5 células y 5 a 8 células. A lo largo del desarrollo embrionario, se pueden definir dos tipos de ciclos: el que corresponde a un ciclo

celular de división, como $ECC2=t4-t2$ y $ECC3=t8-t4$; y la sincronía entre la 2ª y 3ª mitosis, ($s2= t4-t3$) o de la 4ª y 7ª mitosis ($s3=t8-t5$) (Ciray et al., 2014).

Tabla 3. Parámetros morfocinéticos indirectos (Ciray et al., 2014).

Tiempo	Cálculo del evento	Definición de eventos
VP	$tPNf - tPNa$	Duración de los PN
ECC1	$t2 - tPB2$	Duración del primer ciclo celular
ECC2	$t4 - t2$	Duración del segundo ciclo celular
- ECC2a	$t3 - t2$	
- ECC2b	$t4 - t2$	
ECC3	$t8 - t4$	Duración del tercer ciclo celular
- ECC3a	$t5 - t4$	
- ECC3b	$t6 - t4$	
- ECC3c	$t7 - t4$	
- ECC3d	$t8 - t4$	
s2	$t4 - t3$	Sincronización de divisiones celulares
s3	$t8 - t5$	Sincronización del patrón de división
dB	$tB-tSB$	Duración de la blastulación
dexp	$tHN-tE$	Duración de la expansión

4. Reproducibilidad de la evaluación embrionaria

En los años en los que solo existían los criterios morfológicos clásicos, diferentes encuestas realizadas a centros de reproducción asistida en España (Castilla et al., 2010; Hurtado de Mendoza, 2011), sacaron a la luz la gran disparidad de criterios que usaban los diferentes laboratorios para la catalogación embrionaria, disparidad que se agudizaba en la zona intermedia de calidad. La ausencia de criterios comunes afectaba tanto a los parámetros a evaluar, como a los puntos de corte entre categorías, lo que restaba validez a dichos criterios, aumentando la variabilidad entre observadores.

Distintos autores han analizado la variabilidad entre centros o entre centros y expertos en evaluación embrionaria (Arce et al., 2006; Baxter Bendus et al., 2006; Paternot et al., 2009; Ruiz de Assin et al., 2010; Richardson et al., 2015; Storr et al., 2017).

Tanto Arce et al. (2006) como Paternot et al. (2009, 2011) realizaron previo al estudio de fiabilidad y acuerdo, sesiones de formación

Arce et al. (2006) han estudiado la fiabilidad entre centros y expertos en la evaluación de la fragmentación citoplasmática embrionaria, encontrando valores de fiabilidad intermedios. Posteriormente, otros autores han encontrado resultados similares analizando esta misma característica (Paternot et al., 2011). Estos autores también encuentran una disminución en la fiabilidad a la hora de determinar el número de células a medida que aumenta el número de estas. La fiabilidad entre observadores evaluando número de células es contradictoria según las publicaciones (Arce et al., 2006; Paternot et al., 2009, 2011)

En la evaluación de la asimetría Arce et al. (2006) y Paternot et al. (2009) encuentran valores de fiabilidad intermedios entre centros y expertos en el primer caso y entre observadores del mismo centro en el segundo caso, siendo ligeramente superior en día 2 que en día 3. Sin embargo, Paternot et al. (2011) obtuvieron valores de fiabilidad bajos en esta característica entre embriólogos de distintos centros, aunque si observan la misma tendencia comentada anteriormente.

La característica en la que menos fiabilidad existe entre centros y expertos según Arce et al. (2006) es multinucleación, tomando valores moderados en día 2 y bajos en día 3.

Los valores de fiabilidad en la clasificación de embriones entre observadores varían entre intermedia y baja según distintos autores (Arce et al., 2006; Baxter Bendus et al., 2006)

Los autores que han estudiado la fiabilidad en la decisión clínica tomada para embriones obtienen valores óptimos (Paternot et al., 2011).

Storr et al. (2017) analizaron la fiabilidad entre embriólogos de distintos centros en la decisión clínica tomada en embriones en día 5 y en la evolución de la calidad de los mismos. Observan valores de fiabilidad óptimos a la hora de determinar que embriones de día 5 deben transferirse según 5 embriólogos de distintos centros. Pero, observaron valores bajos de fiabilidad tanto en la evaluación de la MCI, del trofoectodermo como de la calidad general del blastocisto. Por el contrario, Richardson et al. (2015) obtuvieron valores sustanciales de fiabilidad tanto en la gradación de embriones en día 5 como en el destino clínico decidido para ellos. En esta última variable encuentran mayor variabilidad entre embriólogos en los embriones que deberían crioconservarse.

Para disminuir esta variabilidad entre centros en la evaluación morfológica embrionaria clásica, se han publicado atlas de embriología (Magli et al., 2012) y diferentes sociedades científicas han recomendado el uso de catalogaciones embrionarias estandarizadas (Racowsky et al., 2010; ALPHA Scientists In Reproductive Medicine and ESHRE Special Interest Group Embryology, 2011; Hurtado de Mendoza et al., 2015). En España, en 2008, se publicó por ASEBIR una catalogación estandarizada de evaluación embrionaria.

Otras recomendaciones para reducir la variabilidad, en catalogación embrionaria, son la asistencia a actividades formativas específicas (Ruiz de

Assin et al., 2011; Lundin et al., 2015) y la participación en CCExt de evaluación embrionaria (Practice Committee of the American Society for Reproductive Medicine et al., 2014; Hurtado de Mendoza et al., 2015; ESHRE Guideline Group on Good Practice in IVF Labs et al., 2016).

Como hemos visto, diferentes autores han estudiado las diferencias entre laboratorios y un grupo de expertos cuando utilizan catalogaciones propias no estandarizadas (Arce et al., 2006; Baxter Bendus et al., 2006; Paternot et al., 2009; Storr et al., 2017). Sin embargo, ningún autor ha analizado esta variabilidad utilizando catalogaciones estandarizadas.

Con la aparición de la tecnología *time-lapse* y su posterior implantación en numerosos centros, existe la necesidad de analizar la reproducibilidad, tanto en la aplicación de los criterios de catalogación embrionaria propuestos por ASEBIR a las imágenes de *time-lapse* como en la determinación de los tiempos de morfocinética. Kirkegaard et al. (2014) demostró la escasa validez externa del modelo de morfocinética descrito por Conaghan et al. (2013). Sugiriendo que esta escasa validez, podría deberse, en parte, a la baja reproducibilidad observada entre usuarios de equipos *time-lapse*. Además, algunos autores han recalcado la necesidad de guías que definan los nuevos eventos que pueden observarse gracias a esta tecnología (Ciray et al., 2014; Kaser y Racowsky, 2014) y una definición exacta del momento de la observación (Lundin et al., 2015).

Aunque actualmente se ha comprobado una baja variabilidad entre usuarios de estos sistemas dentro del mismo centro (Sundvall et al., 2013), se desconoce la variabilidad entre centros a la hora de determinar los tiempos de los distintos eventos del desarrollo embrionario. Por otra parte, los criterios para

evaluar características morfológicas clásica no están bien definidos actualmente para las plataformas *time-lapse* (Hardarson et al., 2002; Chavez et al., 2012). Tampoco se ha analizado la variabilidad entre laboratorios al analizar características morfológicas clásicas utilizando tecnología *time-lapse*. Para estudiar esta variabilidad, como veremos posteriormente, es necesario el desarrollo de programas de CCEExt. Estos programas no solo son necesarios para validar las nuevas tecnologías (Lundin y Ahlstrom, 2015), si no que sin ellos la falta de consenso entre laboratorios conllevaría problemas, como la imposibilidad de generar estudios multicéntricos con una valoración común de la calidad embrionaria, o la dificultad en la interpretación de informes clínicos de otros laboratorios o en la comparación de los datos bibliográficos.

5. Calidad en el laboratorio de reproducción humana asistida

Cualquier prestación sanitaria debe estar centrada en el paciente, esto obliga a los profesionales sanitarios a utilizar las herramientas que tengan a su alcance para asegurar que su actividad satisface las necesidades de los pacientes de una manera segura y eficiente.

5.1. Calidad: Generalidades

El concepto Calidad ha sido definido por diferentes organizaciones. Según la Organización Mundial de la Salud (OMS), que trata de abarcar la perspectiva de los distintos grupos involucrados (pacientes, proveedores, planificadores, sanitarios, etc.); “La Calidad en la Atención en Salud consiste en la apropiada ejecución (de acuerdo con estándares) de intervenciones de probada seguridad, que son económicamente accesibles a la población en cuestión, y que poseen la capacidad de producir un impacto positivo en la mortalidad, morbilidad,

discapacidad y malnutrición”. Por otro lado, la Organización Internacional para la Estandarización (ISO), define la Calidad como el conjunto de caracteres de una entidad (bien producto o servicio) que le confieren una aptitud para satisfacer las necesidades establecidas (las especificadas por un reglamento o los clientes) y las necesidades implícitas (aquellas que no están especificadas pero que conviene identificar y definir).

El concepto de Calidad, al aplicarse a los Servicios de Salud, adoptó modelos provenientes del sector industrial, lo que fortaleció dos elementos que no siempre se tuvieron en cuenta en la atención sanitaria. La importancia del desarrollo de los procesos, entendiendo como proceso el conjunto de actividades mutuamente relacionadas, las cuales transforman elementos de entrada en resultados; y valorar, en los servicios, la satisfacción y la opinión del paciente o usuario.

5.2. Gestión y garantía de la calidad

Dos aspectos importantes de la Calidad son la Gestión de la Calidad y la Garantía de la Calidad.

5.2.1. Gestión de la Calidad

La Gestión de la Calidad se refiere a la cooperación sistemática de toda la organización en interés del producto o servicio para la satisfacción de todos los involucrados, es en sí una cuestión interna de cada organización.

5.2.2. Garantía de la Calidad

Por otro lado, el concepto de Garantía de la Calidad se extiende más allá de la organización, es una evaluación externa basada en acuerdos. Fundamentalmente, es el conjunto de actividades que se llevan a cabo para fijar

normas, vigilar y mejorar el desempeño de tal manera que la atención prestada sea lo más eficaz y segura posible.

5.3. Sistema de calidad y laboratorio

En el sector sanitario existen diferentes modelos de Sistema de Calidad relacionadas con el laboratorio clínico.

En Estados Unidos existe la Ley de Mejora de Laboratorios Clínicos (CLIA), que establece el nivel de calidad para que todos los laboratorios garanticen la exactitud y fiabilidad de los resultados de test de pacientes, con independencia del lugar donde se realicen. La normativa final se basa en la complejidad del método de test; así, cuanto más complicada sea la prueba, más rigurosos serán los requisitos (Ehrmeyer y Laessig, 2001; Rauch y Nichols, 2007).

También en Estados Unidos está la “Joint Commission on Accreditation of Healthcare Organizations” cuya misión es mejorar la calidad de la asistencia proporcionada al público a través de las acreditaciones de asistencia sanitaria.

A nivel europeo, la Fundación Europea para la Gestión de la Calidad (EFQM) establece una serie de normas para la implementación de programas de calidad total (Nabitz et al., 2000).

A nivel internacional, la Organización Internacional de Normalización (ISO) define el sistema de calidad de empresa o de un servicio como el conjunto de la estructura de organización, de responsabilidades, de procedimientos, de procesos y de recursos que se establecen para llevar a cabo la gestión de la calidad.

La familia de normas ISO 9000 son normas de calidad y gestión continua de calidad que se pueden aplicar en cualquier tipo de organización o actividad sistemática, que esté orientada a la producción de bienes o servicios, se componen de estándares y guías relacionados con sistemas de gestión y herramientas específicas.

Derivados de las normas se establecen los requisitos fundamentales del sistema de la calidad en el laboratorio clínico: (a) organización y gestión del laboratorio; (b) instalaciones y equipos; (c) procesos; (d) evaluación y aseguramiento de la calidad; (e) documentación del sistema de la calidad (ISO 15189:2012).

La introducción de un sistema de la calidad en el laboratorio no es gratuita, sino costosa en términos de tiempo y recursos materiales y personales. Cuestiones como el coste-beneficio, profesionalidad y ética, competitividad, etc. deben tenerse en cuenta en el momento de decidir la implantación de un sistema de la calidad. En definitiva, un sistema de la calidad requiere, identificar exactamente los aspectos del laboratorio que hay que cambiar, cual es el procedimiento oportuno para cambiarlos, establecer las motivaciones para que se produzca este cambio y establecer un sistema que haga que el cambio sea permanente cuando decaiga el entusiasmo inicial (ISO 9004:2009).

Los componentes básicos de los Sistemas de Calidad en los laboratorios son: Diseño de calidad, Control de calidad y Mejora de calidad (Figura 16).

Figura 16. Componentes básicos de un Sistema de Garantía de Calidad



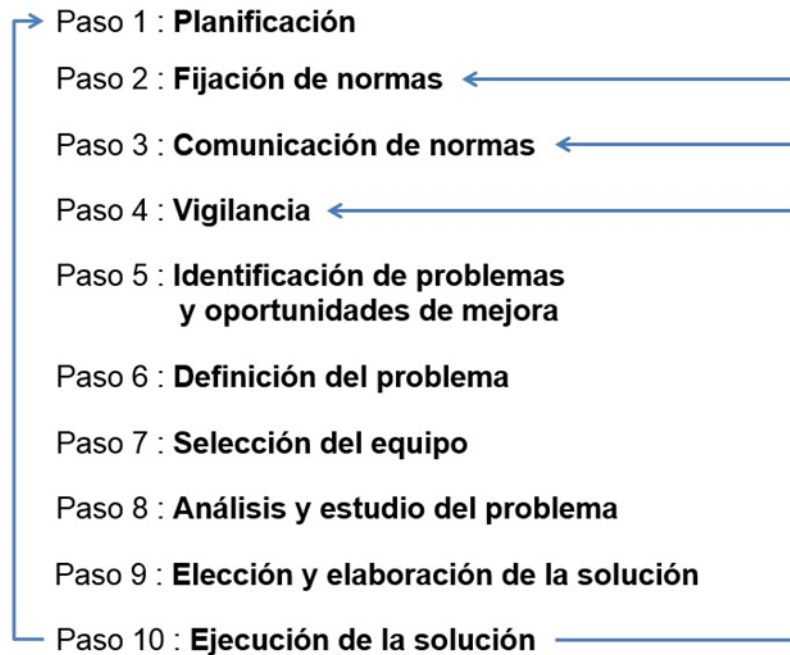
- **Diseño de la Calidad:** planear y desarrollar el proceso. Esto trae como resultado la definición de la misión de la organización, incluyendo sus clientes y servicios, definiendo los medios y recursos y determinando los estándares a aplicar en la prestación del servicio.

- **Control de la Calidad:** seguimiento, supervisión y evaluación que asegure que cada trabajador y cada unidad de trabajo alcance aquellos estándares y consecuentemente brinden servicios de buena Calidad. El control de calidad incluye establecer especificaciones para cada aspecto de un proceso, estableciendo los procedimientos usados en los mismos para determinar la correlación con lo especificado.

- **Mejora de la Calidad:** Promover estándares mediante la resolución continua de problemas y la mejora de procesos.

Los pasos a seguir para el establecimiento de un Sistema de Calidad se representan en la Figura 17.

Figura 17. Pasos a realizar para implementar un Sistema de Calidad



El sistema de la calidad del laboratorio debe describirse en un manual de la calidad, que es un conjunto estructurado de recopilaciones documentales. La estructura general de todos los documentos o procedimientos tendrá un formato lo más homogéneo posible e incluirá apartados tales como tema, objetivos y ámbito de aplicación, responsabilidad afectada por el documento, definiciones y abreviaturas aplicables, actuaciones y método de trabajo, referencias y bibliografía, junto a otra documentación complementaria aplicable. El manual de la calidad debe definir cómo se redacta, evalúa, autoriza y distribuye la documentación de la que consta (ISO 9004:2009).

5.4. Sistema de calidad y laboratorio de reproducción humana asistida

Las nuevas tecnologías, han contribuido a que en los LRHA se logre un producto final, ya sea un informe analítico, rápido, fiable y con alta calidad

científica técnica, o gametos y embriones con una alta viabilidad. Si bien esto es así, no es menos cierto que la calidad conseguida se limita mayoritariamente a la esfera de indicadores técnicos de la calidad: tasa de implantación, tasas de supervivencia o de gestación (Alpha Scientists in Reproductive Medicine, 2012; Castilla y Mantilla, 2016; ESHRE Special Interest Group of Embryology and Alpha Scientists in Reproductive Medicine et al., 2017). Sin embargo, la escasa automatización del laboratorio de reproducción humana asistida (LRHA) (Narayanan et al., 2000), y la alta variabilidad biológica de algunas muestras biológicas analizadas o procesadas (Castilla et al., 2010) deben obligar a establecer un control exhaustivo durante todas las fases de los procedimientos, incluidas las fases pre y posttécnica. Esto permitiría la obtención de resultados clínicamente útiles y el mantenimiento de la seguridad, la eficacia y la eficiencia requeridas por el paciente, es decir, asegurar la ausencia de errores. Además, en el LRHA adquieren mayor relevancia los errores cometidos, no solo por las consecuencias a usuarios, sino a terceros (futuros nacidos).

Los profesionales españoles dedicados al LRHA, han reconocido la necesidad de ampliar los requisitos de la Norma ISO 9001 de sistemas de gestión de la calidad con requisitos específicos de su sector para permitir una gestión más eficiente de la Calidad. Por ello desarrollaron recientemente, junto con la Asociación Española de Normalización y Certificación (AENOR), la Norma UNE 179007:2013 "Sistemas de gestión de la calidad para Laboratorios de reproducción asistida".

Entre otros, los propósitos que se han perseguido para la elaboración de esta Norma han sido los siguientes:

1. Unificar criterios de gestión de calidad del LRHA

2. Ayudar a la planificación del LRHA mediante la gestión por procesos
3. Terminología específica del LRHA
4. Criterios mínimos en cuanto a recursos humanos, infraestructura y requisitos ambientales
5. Criterios mínimos de seguimiento y evaluación de los procesos
6. Facilitar la adaptación a los complejos requisitos legales que afectan al LRHA: trazabilidad, calidad ambiental, destino embrionario, etc.
7. Aportar requisitos mínimos en la comunicación con los clientes y proveedores
8. Criterios mínimos de aseguramiento y control de calidad

Respecto a este último punto, los indicadores de Calidad constituyen una herramienta útil y sencilla para conocer la calidad de un producto, evaluar la conformidad con los requisitos del producto y conocer el rendimiento de un proceso e incluso analizar las tendencias de los mismos (Kohn et al., 2000), siendo un elemento importante para el desarrollo de un sistema de calidad (Mortimer y Mortimer, 2015; ESHRE Guideline Group on Good Practice in IVF Labs et al, 2016)

Existe una gran cantidad de indicadores que podrían darnos información sobre cada proceso del LRHA, pero es conveniente que se lleve a cabo una selección de aquellos que suministren la información más fiable y útil para el LRHA. Para ello, los indicadores pueden basarse en la experiencia previa de nuestro laboratorio, en las recomendaciones de expertos, en los objetivos específicos o generales del LRHA, en un registro de no conformidades (NC),

reclamaciones o encuestas de satisfacción de los pacientes o en oportunidades de mejora del sector. Además, existe una herramienta muy específica para la priorización de estos indicadores denominada Análisis Modal de Fallos y Efectos (AMFE), recientemente aplicada al LRHA en general (Rienzi et al., 2015; Cimadomo et al., 2016; Intra et al., 2016) y a su fase pretécnica en particular (Molina et al., 2017).

5.5. Procesos del laboratorio de reproducción humana asistida

Es importante definir en este punto dos conceptos para su posterior desarrollo: proceso y procedimiento.

Proceso: conjunto de actividades mutuamente relacionadas o que interactúan, las cuales transforman elementos de entrada en resultados. Lo que para un proceso es su producto o resultado, otro lo toma como recurso propio para convertirlo en un nuevo resultado. Transforma elementos de entrada en elementos de salida.

Procedimiento: forma especificada de llevar a cabo una actividad o un proceso. Definen la secuencia de los pasos para ejecutar una tarea.

Un ciclo de tratamiento de reproducción humana asistida engloba varios procesos. Es necesario entender cada uno de estos procesos de manera individual y, además, comprender la relación existente entre ellos.

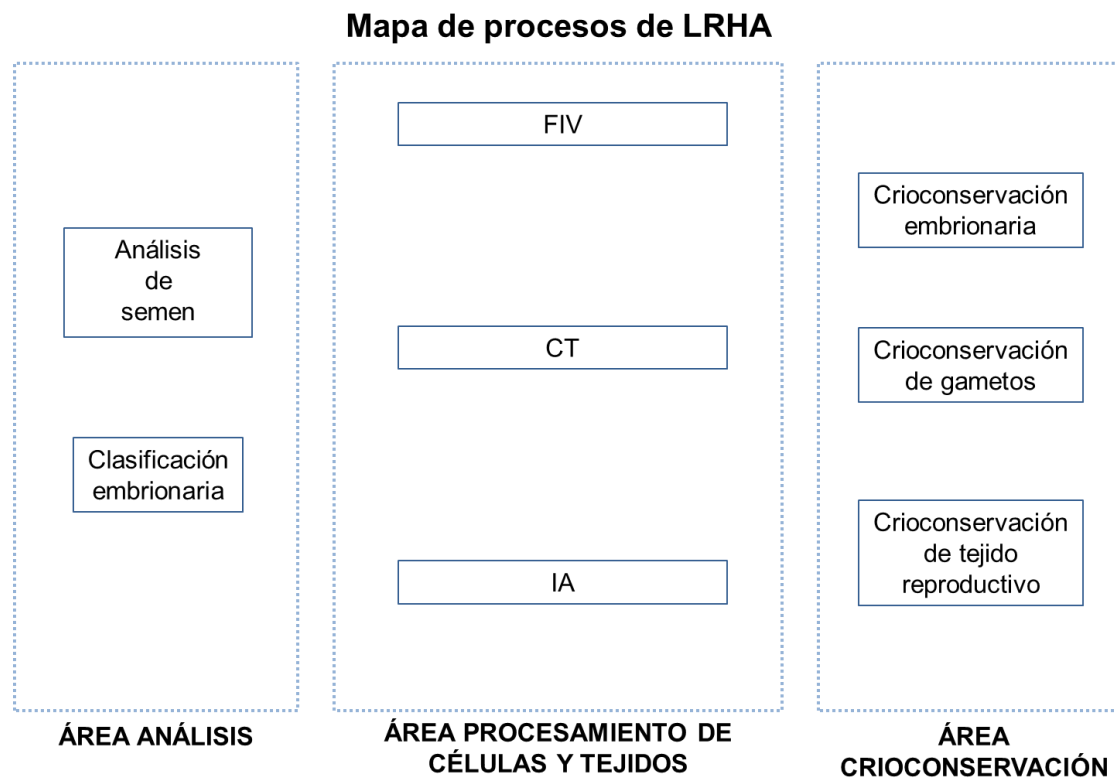
En el LRHA, cada proceso se puede incluir dentro de alguno de las siguientes áreas:

- Análisis: incluye la evaluación de diferentes parámetros seminales y la evaluación de la calidad de ovocitos y embriones.

- Procesamiento de células y tejidos: incluye la manipulación y procesamiento de gametos y embriones para su utilización en las diferentes técnicas de reproducción asistida.
- Crioconservación: incluye la congelación de semen y vitrificación de embriones y ovocitos para utilizarlas en futuras técnicas de reproducción asistida.

La relación entre áreas y procesos de un LRHA se puede representar en la Figura 18.

Figura 18. Mapa de áreas y procesos de LRHA.



Por otro lado, cuando se considera la organización de un laboratorio las actividades a implantar, documentar y realizar pueden clasificarse en: actividades de soporte, actividades de procedimiento y actividades de resultado (Donabedian, 1980):

- El concepto de soporte se usa para englobar recursos humanos, físicos y financieros. Se entiende fácilmente en términos del número apropiado de personal, conveniencia de equipamiento, etc.

- En el contexto del laboratorio, las actividades de procedimiento deben incluir todas las actividades operacionales asociadas con la recogida, transporte, recepción y registro de los especímenes o muestras, su análisis o cultivo, e informe. Las actividades de procedimiento se dividen por tanto en fase pretécnica, fase técnica y fase posttécnica.

- El resultado se usa para indicar un cambio en el estatus habitual y futuro de la salud de un paciente como resultado de una intervención particular, tal como la aplicación de un tratamiento o la realización de una intervención quirúrgica. En los laboratorios de análisis, las actividades de resultado se relacionan con su rendimiento en programas de evaluación externa de la calidad, pero debe abarcar temas más amplios en la auditoría clínica, tales como la evaluación de la elección de la prueba o procedimiento en el contexto del diagnóstico, pronóstico o monitorización de la enfermedad.

A modo de resumen, los procesos en el LRHA se engloban dentro de 3 áreas: análisis, procesamiento de células y tejidos y crioconservación, estando estrechamente relacionados, siendo las salidas de uno, las entradas para otros. Además, cada proceso está compuesto por actividades que pueden corresponder a actividades de soporte, de procedimiento (con 3 fases: pretécnica, técnica o posttécnica), y/o de resultados (Figura 19).

Figura 19. Estructura de un proceso.

Proceso	Actividades	Fases
Denominación del proceso (ej: análisis de semen, FIV, crioconservación de embriones)	Soporte	
	Procedimiento	Pretécnica
		Técnica
		Posttécnica
	Resultados	

5.6. Calidad de las fases de las actividades de procedimiento

5.6.1. Calidad de la fase pretécnica

La fase pretécnica comprende la obtención de la muestra, así como la preparación, el transporte y la conservación de la misma. Para evitar al máximo las disconformidades propias de esta fase, es preciso estandarizar todas las actividades que la integran, y controlar todos los instrumentos propios de esta fase (Young, 1997).

Los factores que influyen en la fase pretécnica pueden ser clasificados en dos grupos, la variabilidad biológica (variaciones biológicas intra e interindividual) y factor técnico, y los dos deberían ser controlados tanto como sea posible:

- Variabilidad biológica
 - Variabilidad biológica intraindividual: Los factores con efecto en parámetros seminales o embrionarios son clasificados como:
 - Fisiológicos: fenómeno en que los valores individuales de los parámetros seminales o embrionarios pueden variar, independientemente del estatus de salud de los pacientes (Álvarez et al., 2003; Fraser, 2001). Pueden ser clasificados como:

- aleatorios (estrés, dieta, ejercicio, abstinencia sexual, estímulo sexual).

- sistemáticos (edad).

- Patológicos: en el caso de semen: fiebre, varicocele, criptorquidia; y en el caso de ovocitos y embriones: endometriosis o síndrome de ovario poliquístico.

- Iatrogénicos: Como resultado de tratamientos médicos o cirugías.

A destacar en este punto la posible influencia de los distintos protocolos de estimulación ovárica en la calidad ovocitaria y embrionaria.

- Variabilidad biológica interindividual: Es el fenómeno en el que los parámetros seminales y embrionarios varían entre sí dentro de una cierta población específica. Por ejemplo, diferencias entre países. Una población puede ser dividida en subconjuntos de población más homogéneos, si se tienen en cuenta factores que aumentan la variabilidad biológica interindividual; así se recomienda que cada laboratorio haga un esfuerzo por tener sus propios valores de referencia para los diferentes parámetros en función de los diferentes grupos.

- Factores técnicos: serán los necesarios para que una muestra para análisis clínico, o una célula o tejido para un uso terapéutico posterior, se obtenga, transporte y conserve en la forma adecuada, del paciente correcto, en el momento correcto y con la adecuada preparación del paciente; transportándose adecuadamente y realizando el análisis o manejo correcto, que responda a la pregunta correcta o necesidad determinada.

La calidad en la fase pretécnica puede garantizarse en el LRHA implantando un sistema de calidad que cumpla los requisitos de las normas ISO 15189:2012, ISO 9001:2015 o UNE 179007:2013.

Actualmente existen dos estrategias complementarias, ya comentadas, para la gestión de la calidad que se aplican a la fase pretécnica. Una de ellas se basa en el AMFE (Giménez et al., 2016), y la otra en el uso de indicadores de Calidad que permiten el análisis de las tasas de NC recopiladas.

La primera se ha aplicado recientemente a la fase pretécnica del LRHA, identificando determinados puntos críticos en esta fase (Molina et al., 2017)

En cuanto a la definición de indicadores de calidad de la fase pretécnica, el grupo de interés de Calidad de ASEBIR ha publicado recientemente un documento donde se definen que requisitos se deben definir en esta fase en los distintos procesos del LRHA (Molina et al., 2017) (Tabla 4).

Tabla 4 Ejemplo de requisitos del Área de análisis del proceso Análisis de semen de la fase pretécnica del LRHA (Molina et al., 2017).

ACTIVIDAD							
ANÁLISIS DE SEMEN	Solicitud*	Preparación del paciente	Obtención	Identificación	Transporte	Manipulación	Conservación de la muestra
		<ul style="list-style-type: none"> - Identificación del paciente/a: nombre, edad etc. - Tipo de petición: urgente, normal. - Datos de filiación del paciente/a. - Datos administrativos de la solicitud. - Pruebas o estudios solicitados. 	<ul style="list-style-type: none"> - Tiempo de abstinencia. - Medidas higiénicas para recoger la muestra. - Posibles interferencias: estados febriles, fármacos, drogas, etc. - Estudio serológico realizado. 	<ul style="list-style-type: none"> -Método de obtención del eyaculado. - Tipo de material - Características del espacio destinado a la obtención del semen. - Tipo de recipiente. 	<ul style="list-style-type: none"> - Datos del paciente: nombre y apellidos, número de historia, teléfono, dirección. - Fecha desde la última eyaculación. - Hora de obtención/ Hora de recepción. - Muestra completa. - Responsable de la identificación de la muestra. - Código de identificación. 	<ul style="list-style-type: none"> - Tipo de material homologado y recomendado para el transporte de semen. - Temperatura óptima de transporte. - Tiempo máximo de entrega. 	<ul style="list-style-type: none"> -Características físicas de los espacios destinados a la manipulación de semen. - Características ambientales para la manipulación de semen. - Medios y material usados en las distintas fases de manipulación. - Intervalos de tiempo óptimos para la manipulación del semen.

* Ante una solicitud verbal de ampliación del estudio, siempre debe existir la autorización de un facultativo del área correspondiente del laboratorio y registrarse la prueba solicitada en la petición previa, indicando la fecha de ampliación, así como el nombre del clínico que la solicita.

Para la redacción de estos requisitos debe utilizarse un lenguaje claro, conciso y de fácil comprensión. Los documentos, además de incluir los requisitos en una secuencia lógica, deben explicar las razones de cada epígrafe con objeto de favorecer su cumplimiento y antes de difundir el documento en la población usuaria, es recomendable comprobar su eficacia mediante ensayos de campo.

Al establecer los requisitos comentados de la fase pretécnica del LRHA, se ha indicado cómo se deben desempeñar las actividades en esta fase. Si existe una desviación que no resulta conforme a lo que se describió, este se define como NC en nuestro sistema.

Cuando ocurra una NC, incluida la derivada de una queja o reclamación del cliente, se recomienda emprender acciones para controlarla de manera inmediata (ISO 9001:2015 o UNE 179007:2013). Posteriormente, eliminar las causas que la originaron, revisar la eficacia de las acciones y si fuera necesario hacer los cambios pertinentes en las actividades relacionadas en nuestro laboratorio.

Para poder gestionar las NC de este modo es preciso que el laboratorio disponga de un registro eficaz de las mismas. El registro de NC es clave para la utilización de indicadores que permitan el seguimiento y monitorización del cumplimiento de requisitos.

Para la monitorización del cumplimiento de estos requisitos del comentado documento de ASEBIR se han descrito indicadores de calidad para la fase pretécnica de los procesos del LRHA (Molina et al., 2017). En la Tabla 5 se presenta un ejemplo de estos. Por su reciente implantación todavía no se disponen de las especificaciones de los indicadores de la fase pretécnica.

Tabla 5. Indicadores de calidad para Actividades de la Fase pretécnica del proceso de Análisis de semen del Área de Análisis del LRHA (Molina et al., 2017).

	NO CONFORMIDAD	DEFINICIÓN DE LA NO CONFORMIDAD	“INDICADOR PREANALÍTICO PROPUESTO”
Solicitud para preparación de muestra de semen	Solicitud inadecuada y/o incompleta.	La solicitud presenta parámetros seminales no necesarios, faltan datos y/o existen dos o más solicitudes con igual código.	Porcentaje de “Número total de solicitudes inadecuadas e incompletas / número total de solicitudes de análisis de semen”.
Preparación del paciente para la recogida de semen	No ha seguido la información de recogida.	La muestra no se ha recogido en el intervalo de tiempo establecido.	Porcentaje de “Número de muestras que no se han recogido en el intervalo de tiempo establecido/ número total de análisis de semen”.
	Ausencia de analíticas de ETS (serología).	El paciente no tiene las serologías de enfermedades de transmisión sexual.	Porcentaje de “Número de muestras de pacientes sin resultados serológicos de ETS/ número total de análisis de semen”.
	Paciente ausente.	El paciente no se personaliza para recoger la muestra.	Porcentaje de “Número de pacientes que no acuden a recoger la muestra para su preparación /número total de pacientes citados para análisis de semen”.
	Hoja de recogida de muestra de semen incompleta.	Paciente que no completa la hoja de información (fiebre, medicación, viajes en los últimos meses).	Porcentaje de “Número de hojas de recogida incompletas/ número total de análisis de semen”.
Obtención de la muestra de semen	Material inadecuado de recogida de muestra de semen.	Frasco para la obtención no adecuado al protocolo normalizado de trabajo.	Porcentaje de “Número de muestras recogidas en el contenedor inadecuado/ número total muestras recogidas para análisis de semen”.
	Muestra incompleta.	Pérdida de parte de la muestra en el proceso de recogida.	Porcentaje de “Número de muestras incompletas en el contenedor/ número total de muestras recogidas para análisis de semen”.

Tabla 5. Indicadores de calidad para Actividades de la Fase pretécnica del proceso de Análisis de semen del Área de Análisis del LRHA (Molina et al., 2017) (Continuación).

	Muestra no recogida.	El paciente no puede recoger la muestra.	Porcentaje de “Número de muestras no recogidas/número total de pacientes que intentan recoger la muestra para análisis de semen”.
	Obtención inadecuada de la muestra.	La muestra no se obtiene por masturbación.	Porcentaje de “Número de muestras no recogidas por masturbación/número total de muestras recogidas para análisis de semen”.
	Muestra de semen contaminada.	La muestra de semen está contaminada con orina y/o gérmenes.	Porcentaje de “Número de muestras contaminadas/número total de muestras recogidas para análisis de semen”.
	Recogida en la sala inadecuada.	El paciente no recoge la muestra en las salas de recogida establecidas.	Porcentaje de “Número de muestras recogidas en salas inadecuadas/ número total muestras recogidas para análisis de semen”.
	Necesidad de medicación para recoger la muestra de semen.	El paciente requiere medicación para obtener la muestra de semen.	Porcentaje de “Número de muestras obtenidas con ayuda de medicación/ número total de muestras recogidas para análisis de semen”.
Identificación de la muestra de semen	Muestra sin identificar, incompleta ó mal identificada.	Muestra sin nombre y/o código, datos ilegibles y/o incompletos, código y/o nombres incompletos.	Porcentaje de “Número total de muestras incorrectamente identificadas/ número total de muestras recogidas para análisis de semen”.

Tabla 5. Indicadores de calidad para Actividades de la Fase pretécnica del proceso de Análisis de semen del Área de Análisis del LRHA (Molina et al., 2017) (Continuación).

Transporte de la muestra de semen	Transporte inadecuado de la muestra.	No hay cumplimiento de las normas para permitir la calidad de la muestra.	Porcentaje de “Número de muestras que no cumplen las condiciones establecidas de transporte /número total de muestras recogidas para análisis de semen”.
	Demora en la entrega.	Tiempo desde la recogida de la muestra hasta su entrega en el laboratorio excedido.	Porcentaje de “Número de muestras transportadas en tiempo inapropiado/número total de muestras recogidas para análisis de semen”.
Manipulación	Temperatura inadecuada, espacio inadecuado, insuficiente y/o contaminado	La temperatura del laboratorio no cumple con las condiciones establecidas. Espacio físico insuficiente y/o no hay división de espacio para muestras de infecciosos y/o espacio destinado a la manipulación con gérmenes.	Porcentaje de “Número total de muestras preparadas en condiciones inadecuadas/ número total de muestras recogidas para análisis de semen”.
	Ausencia de fungible, medios, material inadecuado y/o manipulación incorrecta.	El fungible no está preparado para el análisis de la muestra según describe el protocolo y/o so de materiales no homologados y/o no óptimos, uso de medios y/o técnicas no óptimas para su manipulación.	Porcentaje de “Número total de muestras preparadas con material inadecuado y manipulación incorrecta/ número total de muestras preparadas para análisis de semen”.
Conservación de la muestra de semen	Tiempo medio para el análisis de la muestra inadecuado.	Tiempo desde la recogida de la muestra hasta su análisis excedido.	Porcentaje de “Número de muestras analizadas en tiempo inadecuado/número total de muestras analizadas”.
	Conservación incorrecta de la muestra.	Uso de medios y/o técnicas no óptimas para su conservación.	Porcentaje de “Número de muestras que no cumplen las condiciones establecidas de conservación/número total de muestras recogidas para análisis de semen”.

5.6.2. Calidad de la fase técnica

La Calidad en esta fase, comprende los pasos prácticos encaminados a asegurar que los errores de medida cometidos o resultados negativos obtenidos al mantener la viabilidad de gametos y embriones estén dentro de un margen aceptable. Esto significa que los errores cometidos y resultados negativos obtenidos inevitables son cuantificados para permitirnos decidir si son de una magnitud aceptable y que errores y resultados negativos inaceptables son descubiertos para decidir qué acción correctiva puede ser tomada.

5.6.2.1. Fase técnica de las actividades de procesamiento del área de análisis

El valor verdadero de una magnitud se obtendría con una medición perfecta, pero como tal medición no existe, el valor verdadero es incognoscible. Por lo tanto, aunque en la teoría se use el valor verdadero, en la práctica se utilizan otros valores considerados convencionalmente como verdaderos.

La variabilidad analítica puede desglosarse en aleatoria y sistemática, según se trate de errores sistemáticos o aleatorios. El error aleatorio es la diferencia entre el resultado de una medición y la media que se obtendría de un número infinito de mediciones (en la práctica vale con $n \geq 20$) de la misma magnitud particular (Stöckl et al., 1996). El error sistemático es la diferencia entre la media que se obtendría de un número infinito de mediciones (en la práctica vale con $n \geq 20$) del mismo mesurado y su valor verdadero. El error total (o inexactitud) es el concepto que integra la variabilidad aleatoria y la sistemática de manera que es la diferencia entre el resultado de una única medición y el valor verdadero del mesurado.

La variabilidad analítica debida a fluctuaciones aleatorias se cuantifica mediante la desviación típica, varianza, o coeficiente de variación. Según la definición de la Federación Internacional de Química Clínica, la “imprecisión” es la desviación típica o el coeficiente de variación de los resultados de un conjunto de medidas repetidas. Idealmente la imprecisión interdiaria de cualquier procedimiento de medida debería ser nula. Es importante establecer cuál es la desviación típica interdiaria máxima tolerable, de manera que esta debe ser tal que permita diferenciar entre una variación intraindividual fisiológica y una variación intraindividual debida a algún estado patológico.

5.6.2.2. Fase técnica de las actividades de procesamiento de las áreas de procesamiento de células y tejidos y crioconservación

En la fase técnica del procesamiento de células y tejidos y crioconservación, también nos podemos encontrar con resultados negativos debido a errores técnicos (temperaturas de placas calefactadas e incubadoras, concentración de CO₂, etc.) o los causados por sustancias exógenas (fármacos y otros xenobióticos) o por sustancias endógenas a concentraciones elevadas o contaminación.

Como se ha comentado para la fase pretécnica, una manera de identificar y gestionar la Calidad en el LRHA es el uso de indicadores de calidad. Pero al contrario de lo que sucede en la fase pretécnica si se han descrito especificaciones de estos indicadores de esta fase técnica que son actualizadas anualmente por ASEBIR (López-Regalado y Castilla, 2017).

5.6.2.3. Indicadores y especificaciones

Los estándares o especificaciones numéricas de los indicadores de calidad (también llamados objetivos de la calidad, estándares deseables, y objetivos de la prestación) son requeridos en el laboratorio clínico en estudios de evaluación, para decidir el cumplimiento de las prestaciones de un método o técnica, para ayudar a los organizadores de programas de evaluación externa y pruebas de actitud en el diseño y ejecución adecuada de los programas, y para ayudar a los fabricantes de instrumentos y reactivos en el diseño, producción y estudios de mercado. Lo más importante de todo ello es que los laboratorios aseguren que la calidad alcanzada es la adecuada para el cuidado de los pacientes.

Estas especificaciones de calidad son necesarias en el laboratorio clínico, pues éste solamente debe usar procedimientos validados para confirmar que los procedimientos son adecuados para el uso previsto (ISO 15189:2012). Las validaciones deben ser tan exhaustivas como sea necesario para satisfacer las necesidades del uso dado o el campo de utilización. Además, las especificaciones de calidad para cada procedimiento usado deben estar relacionadas con el uso previsto de ese procedimiento.

Existían muchas recomendaciones publicadas para calcular las especificaciones de calidad técnica y es difícil para los no expertos decidir el modelo adecuado, constituyendo esto un problema para seleccionar la especificación más apropiada para su uso en la planificación de la calidad.

Estas recomendaciones, fueron planteadas y recogidas en una clasificación que atiende a un modelo jerárquico, en el consenso final alcanzado en la

conferencia de Estocolmo en 1999 (Kenny et al., 1999) y revisado recientemente en una reunión celebrada en Milán (Sandberg et al., 2015), donde se promovieron acuerdos en la definición de la estrategia adecuada para establecer especificaciones globales de la calidad, que puedan aceptarse en todo el mundo, prescindiendo de que el laboratorio sea grande o pequeño, público o privado, desarrollado o en desarrollo.

Los criterios alcanzados para determinar las especificaciones de calidad fueron: Basados en su impacto clínico, basados en las variaciones biológicas o derivadas de la opinión de los clínicos, basados en recomendaciones profesionales, basados en los resultados de las evaluaciones externas de la calidad y basados en el estado del arte.

En el caso concreto del laboratorio de reproducción, existe literatura sobre las especificaciones de calidad de indicadores de procedimiento (fase técnica) del análisis de parámetros seminales, las cuales han sido calculadas basadas en la variación biológica (Álvarez et al., 2003), basadas en el estado del arte (Castilla et al., 2005), y basadas en la opinión de los clínicos (Aguilar et al., 2008). Estas especificaciones de calidad difieren entre métodos, lo que coincide con lo observado anteriormente para otras magnitudes biológicas empleando distintos modelos de establecimiento de especificaciones de calidad analítica (Lott, 1999; Petersen et al., 2001).

En cuanto a los indicadores de calidad de las actividades de procedimiento (fase técnica) del área de procesamiento de gametos y embriones y crioconservación, se han publicado recientemente por parte de la Sociedad Europea

de Reproducción Humana y Embriología (ESHRE) y Alpha tras una reunión de consenso (ESHRE Special Interest Group of Embryology and Alpha Scientists in Reproductive Medicine, 2017). Hace diez años, en 2008, ASEBIR elaboró un documento de consenso sobre “Estandarización de Indicadores en el Laboratorio de Reproducción Asistida” elaborado por su Comisión de Calidad (Santos et al., 2007), y en 2009 elaboró por primera vez estándares para algunos de estos indicadores que han sido anualmente revisados según el estado del arte (Castilla y Mantilla, 2016). De igual modo, la Norma UNE 197007 elaborada, entre otros, por ASEBIR y AENOR, incluye un listado de indicadores para el Laboratorio de Reproducción (Tabla 6), muchos de ellos coincidentes con los ya comentados, elaborados por ASEBIR.

Tabla 6. Lista de indicadores de calidad del laboratorio de reproducción asistida humana (UNE 179007:2013).

INDICADOR
1. Porcentaje de fecundación normal FIV/ICSI en ciclos con ovocitos propios en fresco.
2. Porcentaje de fecundación normal FIV/ICSI en ciclos con ovocitos donados en fresco.
3. Porcentaje de gestación clínica por transferencia en ciclos de FIV/ICSI con ovocitos propios en fresco.
4. Porcentaje de gestación clínica por transferencia en ciclos de FIV/ICSI con ovocitos propios crioconservados.
5. Porcentaje de gestación clínica por transferencia en ciclos de FIV/ICSI con ovocitos donados en fresco.
6. Porcentaje de gestación clínica por transferencia en ciclos de FIV/ICSI con ovocitos donados crioconservados.
7. Tasa de implantación en ciclos de FIV/ICSI con ovocitos propios en fresco.
8. Tasa de implantación en ciclos de FIV/ICSI con ovocitos donados en fresco
9. Tasa de implantación en ciclos de embriones crioconservados con ovocitos propios.
10. Tasa de implantación en ciclos de embriones crioconservados con ovocitos donados.

Tabla 6. Lista de indicadores de calidad del laboratorio de reproducción asistida humana (UNE 179007:2013) (Continuación).

11. Porcentaje de gestación clínica por transferencia de embriones crioconservados procedentes de ovocitos propios.
12. Porcentaje de gestación clínica por transferencia de embriones crioconservados procedentes de ovocitos donados.
13. Número de embriones transferidos por embarazo en ciclos de FIV/ICSI con ovocitos propios en fresco.
14. Número de embriones transferidos por embarazo en ciclos de FIV/ICSI con ovocitos donados en fresco.
15. Número de embriones transferidos por embarazo en ciclos de criotransferencias con ovocitos propios.
16. Número de embriones transferidos por embarazo en ciclos de criotransferencias con ovocitos donados.
17. Porcentaje de embriones utilizados en ciclos de FIV/ICSI con ovocitos propios en fresco.
18. Porcentaje de embriones utilizados en ciclos de FIV/ICSI con ovocitos donados en fresco.
19. Porcentaje de supervivencia ovocitaria tras descricriocervación en ciclos con ovocitos propios.
20. Porcentaje de supervivencia ovocitaria tras descricriocervación en ciclos con ovocitos donados.
21. Porcentaje de supervivencia embrionaria tras descricriocervación en ciclos con ovocitos propios.
22. Porcentaje de supervivencia embrionaria tras descricriocervación en ciclos con ovocitos donados.
23. Porcentaje de gestación clínica en ciclos de inseminación artificial conyugal.
24. Porcentaje de gestación clínica en ciclos de inseminación artificial de donante.
25. Tasa de gestación múltiple en ciclos de FIV/ICSI con ovocitos propios.
26. Tasa de gestación múltiple en ciclos de FIV/ICSI con ovocitos donados.
27. Tasa de gestación múltiple en ciclos de inseminación artificial conyugal.
28. Tasa de gestación múltiple en ciclos de inseminación artificial de donante.
29. Número medio de embriones por transferencia en ciclos de FIV/ICSI con ovocitos propios en fresco.
30. Número medio de embriones por transferencia en ciclos de FIV/ICSI con ovocitos donados en fresco.
31. Porcentaje de embriones crioconservados por ciclo con ovocitos propios en fresco.
32. Porcentaje de embriones crioconservados por ciclo con ovocitos propios donados.

Se han utilizado dos métodos para calcular las especificaciones de estos indicadores: el método basado en el estado del arte y el consenso de expertos.

El uso del estado del arte ha sido propuesto por muchos grupos, entre ellos la Sociedad Francesa de Biología Clínica (Vassault et al., 1999) y la Asociación Española de Farmacéuticos Analistas (AEFA) (Calafell et al., 2002). El estado del arte como se evidencia de las Pruebas de Aptitud o de los Programas de Evaluación Externa de la Calidad ha sido defendido ampliamente como especificación de la calidad, particularmente cuando las prestaciones alcanzadas por los mejores laboratorios, típicamente el mejor 25%, se toma como valor diana. El estado del arte hace referencia al nivel más alto de desarrollo conseguido en un momento dado sobre cualquier aparato, técnica o campo científico. El concepto subyacente es que, si uno de cada cuatro laboratorios puede conseguir este nivel de calidad, entonces la tecnología y metodología es suficiente para que todos los laboratorios puedan conseguir el mismo nivel de prestación analítica.

En el campo de la embriología clínica se puede aplicar este concepto partiendo de los registros de actividad de técnicas de reproducción asistida, considerando como especificación aquellos resultados obtenidos por los mejores laboratorios.

Los criterios usados por ASEBIR, para establecer las especificaciones de los indicadores de la UNE 179007:2013, fueron los límites utilizados por AEFA para el cálculo de las especificaciones de calidad analíticas en la que se definen 3 niveles de especificaciones:

- Mínima: aquella que es capaz de conseguir el 95% de los laboratorios

- Deseable: aquella que es capaz de conseguir el 75% de los laboratorios
- Óptima: Aquella que es capaz de conseguir el 25% de los laboratorios

5.6.3. Calidad de la fase posttécnica

Comprende todas las actividades, desde la obtención del resultado de la medición de una magnitud, del procesamiento de células y tejidos o de la crioconservación, hasta su llegada al solicitante.

Los diferentes sistemas de gestión de la calidad recomiendan a los laboratorios, centrar sus esfuerzos de mejora, no sólo en la fase técnica, sino también en el resto de fases de cualquier actividad de procedimiento (ISO 15189:2012).

La escasa comunicación entre los profesionales del laboratorio y los usuarios de este afecta a la calidad de las fases pre y posttécnica. Concretamente, la comunicación posttécnica implica a la comunicación entre los profesionales del laboratorio y clínicos, puntualidad de la información, notificación de resultados de pruebas anormales o la presentación de información relevante a través de informes y comentarios interpretativos. Fallos en la comunicación, conducen a errores, que afectan a la seguridad del paciente y al uso ineficiente e ineficaz de los recursos.

El informe de los resultados ha de ser claro y simple, y debe incluir la determinación o actividad realizada, y en el caso de procedimientos analíticos, la unidad de medida y el intervalo de referencia según sexo y edad. Los laboratorios deben tener personal que pueda asesorar a los clínicos en la evaluación e

interpretación de los resultados de las pruebas y técnicas del laboratorio, y de la precisión y exactitud de los procesos de análisis y de la eficacia del procesamiento de células y tejidos o de la crioconservación. El término “elaboración de informes” incluye tanto el contenido del informe y la validación de los datos, como la información en si misma.

En cuanto a los límites de referencia, cuando se dispone de varios valores de referencia de un mismo individuo obtenidos en distintos momentos, se puede estimar el intervalo de referencia individual. Así, los valores de referencia individuales deben haber sido obtenidos con procedimientos de medida que produzcan resultados intercambiables: no pueden efectuarse comparaciones ni pueden establecerse intervalos de referencia individual mezclando resultados que, pese a estar obtenidos en un mismo individuo, no se hayan producido con el mismo error sistemático y con la misma imprecisión (Harris y Boyd, 1995).

Por otra parte, los límites de referencia poblacionales de cualquier magnitud analítica se obtienen a partir de valores de referencia producidos con una imprecisión determinada. Por lo tanto, los valores de referencia obtenidos en una misma muestra de referencia por dos procedimientos de medida que poseen distinta imprecisión son diferentes, y también serán diferentes los límites de referencia.

Los indicadores de calidad son una buena herramienta para monitorizar todas las fases de una actividad, como se ha comentado anteriormente, existe bastante literatura sobre indicadores de las fases pretécnica y técnica. Sin embargo, el desarrollo de indicadores para la fase posttécnica ha sido más lento (Sciacovelli

et al., 2017). Los principales indicadores de calidad usados en la fase posttécnica son los relacionados con informes incorrectos o con un tiempo inadecuado de respuesta (Sciacovelli et al., 2017).

5.7. Control de calidad

Como se ha comentado, la variabilidad analítica es un fenómeno propio de los procesos de medida, que ocasiona que los resultados obtenidos al medir repetidamente una magnitud en un mismo espécimen puedan ser ligeramente diferentes entre sí (Haeckel, 1993). Este concepto, aplicado al manejo de gametos o al cultivo embrionario en el LRHA, es el que hace que los resultados obtenidos al analizar, procesar o crioconservar alícuotas de un mismo espécimen sean ligeramente diferentes entre sí. Todo esto nos obliga a realizar un conjunto de acciones internas o externas conocido como “control de la calidad”. Estas acciones deben estar planificadas, realizarse sistemáticamente y ser demostrables.

5.7.1. Control de calidad interno

El control de calidad interno (CCInt) analítico, pretende en primera instancia, detectar la existencia de errores de medida superiores a los tolerables, teniendo en cuenta las características analíticas del procedimiento de medida empleado. Existen para ello diferentes métodos como gráficos de Levey-Jennings, método de las sumas acumuladas, control con especímenes de pacientes, control mediante la media de los resultados de los pacientes, control mediante duplicados interseriales y el análisis de materiales de referencia.

En cuanto al CCInt en el cultivo de gametos y embriones, debemos diferenciar dos aspectos. En primer lugar, los controles diarios a realizar en los equipos. Un mal desempeño de este control podría identificarse varias semanas después cuando la tasa de embarazo disminuya, retrasando la implementación de acciones correctivas (Elder y Kastrop, 2003). Este CCInt, consiste en la comprobación de las temperaturas de incubadoras, estufas y placas calefactadas, sin confiar en las pantallas digitales, medición diaria de la concentración de CO₂ en incubadoras y tanques de presión de CO₂, revisión de los niveles de agua en los aparatos que así lo precisen, revisión de temperatura en congeladores y frigoríficos y revisión de esta y de los niveles de nitrógeno líquido (N₂) en los tanques de almacenamiento (Elder y Kastrop, 2003).

El otro aspecto de CCInt es el registro de los resultados de fecundación, desarrollo de embriones, supervivencia embrionaria y ovocitaria tras su descricioconservación, y su comparación con registros históricos del mismo laboratorio, lo que evidencia a corto plazo la buena calidad del procedimiento.

Un factor clave en el mantenimiento de la calidad, es la correcta formación de los nuevos miembros del equipo. El término de “curva de aprendizaje” comprende el esfuerzo requerido para adquirir una nueva habilidad en un determinado periodo de tiempo. El tipo de curva (rápida o gradual) dependerá de la dificultad técnica inherente y de la heterogeneidad de los casos. Así, por ejemplo, la preparación del semen es técnicamente fácil, pero el tener que individualizar la técnica a cada muestra hace que para obtener óptimos resultados sea necesaria una curva de aprendizaje gradual. Sin embargo, la microinyección de espermatozoides es

técnicamente difícil, pero generalmente se aplica el mismo protocolo a todas las parejas.

Un sistema de aprendizaje estandarizado es esencial para evitar futuros problemas. El proceso de instrucción tiene que ser seguido de cerca, documentado y aprobado por el director del laboratorio. Por lo tanto, es importante establecer los criterios de la instrucción y los umbrales que deben alcanzarse antes de comenzar a tratar muestras de pacientes sin supervisión. Igualmente, es deseable aplicar las medidas necesarias para que, una vez alcanzados esos umbrales, estos sean mantenidos. Se han descrito diferentes niveles en la formación del personal del laboratorio de reproducción: comprensión básica de las técnicas, puede trabajar bajo supervisión, puede trabajar sin supervisión y puede entrenar a otras personas (Keck et al., 2005; Hurtado de Mendoza et al., 2008).

Asociado a esto, una parte importante del control interno es controlar las diferencias interobservador y realizar reuniones intracentro para homogenizar criterios, para que las diferencias entre los distintos técnicos sean mínimas. En la Tabla 7 se muestran distintos ejemplos para controlar estas diferencias. ASEBIR desde 2015, pone a disposición de los laboratorios de embriología españoles imágenes de varios embriones para ser evaluados por miembros de ese mismo laboratorio con el objetivo de fomentar la implantación de programas de CCInt de evaluación morfológica embrionaria (ASEBIR, 2017)

Tabla 7. Ejemplos de control de calidad interno para actividades de procedimiento (Fase Técnica) (Castilla y Mantilla, 2016).

Área	Proceso	Actividad	Control de calidad interno
Análisis	Análisis de semen		Semen crioconservado
		Determinación de la concentración espermática	Suspensión bolitas látex Suspensión espermática Videos
		Determinación de la movilidad espermática	Semen crioconservado Videos
		Determinación de la morfología espermática	Portaobjetos teñidos y/o fijados Semen crioconservado Videos y/o fotos
		Determinación de la vitalidad espermática	Portaobjetos teñidos Semen crioconservado Videos
		Realización un test hipoosmótico	Semen crioconservado Videos
		Determinación de anticuerpos anti-espermatozoide directa	Semen crioconservado Videos
		Determinación de anticuerpos anti-espermatozoide indirecta	Sueros
		Recuperación de espermatozoides móviles	Realización de REM con el mismo donante conocido
			Clasificación embrionaria

Tabla 7. Ejemplos de control de calidad interno para actividades de procedimiento (Fase Técnica) (Castilla y Mantilla, 2016) (Continuación).

Área	Proceso	Actividad	Control de calidad interno
Procesamiento de células y tejidos	FIV/ICSI CT IA	Determinación de la toxicidad de fungible	Ensayos de citotoxicidad periódicos
			Registro histórico de tasas de gestación
Crioconservación	Crioconservación de gametos (masculino)		Tests de descongelación de un mismo eyaculado congelado en distintas alícuotas por diferente personal
	Crioconservación de gametos (femenino)		Tasa de supervivencia de ovocitos no fecundados crioconservados por distintos embriólogos
	Crioconservación de embriones		Tasa de supervivencia embrionaria por embriólogo

5.7.2. Control de calidad externo

Los programas de evaluación externa de la calidad más conocidos en los laboratorios clínicos son los basados en la participación de diversos laboratorios que miden, con los mismos o diferentes procedimientos de medida, una o más magnitudes en los mismos materiales de control, y en el que los resultados obtenidos por un laboratorio son comparados con los resultados obtenidos por el resto de los laboratorios participantes. Por tanto, permite conocer cómo están funcionando los procedimientos de un determinado laboratorio en relación a cómo

funcionan en otros laboratorios (OMS, 2010). De otra parte, los programas de CCExt constituyen la base de la transferibilidad de resultados entre laboratorios.

Los programas de evaluación externa de la calidad suelen estar organizados por fabricantes de materiales de control, organismos oficiales u organizaciones científicas. La participación suele ser voluntaria, aunque en algunos países la ley obliga a los laboratorios a participar en un programa de evaluación externa de la calidad y sus resultados pueden ser inspeccionados oficialmente.

Cada programa de evaluación externa de la calidad establece el número de materiales de control, su frecuencia de distribución y los valores de las magnitudes cuyos procedimientos de medida se desea evaluar, dependiendo todo ello de la finalidad concreta para la que haya sido diseñado.

Además de las características generales sobre la estructura del programa de evaluación externa de la calidad, los participantes reciben información sobre el origen de los materiales de control e instrucciones claras sobre su manipulación, estabilidad, condiciones de almacenamiento y fecha límite de entrega de los resultados (IUPAC, 1996).

Para que un programa de evaluación externa de la calidad sea fiable deben cumplirse los siguientes puntos (Federación Internacional de Química Clínica, 1983):

1. Los distintos materiales de control de un mismo lote deben ser totalmente homogéneos entre sí y su estabilidad debe mantenerse a lo largo de todo el periodo de control.

2. El número de laboratorios participantes debe ser el mayor posible y la diversidad de los métodos de medida de una misma magnitud también debe ser la mayor posible, con lo que la estimación de los valores consensuados, tanto por método como global, será fiable y permitirá la comparación entre laboratorios que utilizan el mismo método, y una buena estimación del error sistemático.

3. Deben utilizarse materiales de control con diferentes valores de las magnitudes en estudio, intentando abarcar todo el intervalo de medición, o al menos, con valores cercanos a las decisiones clínicas.

Existen dos tipos principales de programas de evaluación externa de la calidad analítica: tipo I y tipo II. Las diferencias principales entre ambos tipos son la frecuencia con que se realizan las mediciones en los materiales de control y el hecho de que los valores de esos materiales sean conocidos por los laboratorios participantes o no lo sean (Federación Internacional de Química Clínica, 1983).

- **Tipo I:** En estos programas se utilizan los resultados obtenidos de forma esporádica, aunque periódica, por los laboratorios participantes. Se utilizan diversos materiales de control cuyos valores no son conocidos por los laboratorios participantes. Durante los 6 ó 12 meses que suele durar cada ciclo de control, cada laboratorio mide en cada material de control con una cadencia quincenal, mensual o trimestral las magnitudes correspondientes a los procedimientos que se desea controlar y, finalmente, envía los resultados al organizador. Poco tiempo después, el organizador envía a cada laboratorio participante, para cada procedimiento de medida sometido a control, un informe en el que se comparan los resultados obtenidos por el participante con los obtenidos por los laboratorios que utilizan el

mismo método de medida y con los obtenidos por todos los laboratorios participantes, sea cual sea el método con el que miden una misma magnitud. Finalizado el periodo de control, el organizador elabora un informe global en el que evalúa la calidad analítica obtenida por el participante para cada procedimiento y la compara con la obtenida por el resto de los laboratorios participantes. En algunos programas, en el informe correspondiente al final del periodo de control, se indica la clasificación basada en la imprecisión interdiaria y el error sistemático, de cada uno de los procedimientos utilizados por el participante en relación con el resto de los participantes (IUPAC, 1996).

Tipo II: En estos programas se utilizan los resultados obtenidos cotidianamente en el CCIInt. En estos programas, cuya duración suele ser de un año o más, los materiales de control pueden estar valorados o no, aunque si no lo están, cada laboratorio participante les asigna sus propios valores. Así pues, los materiales de control se usan cada vez que se realiza una serie de mediciones para controlar los resultados observados en los pacientes. Los resultados de control obtenidos en el CCIInt se envían mensualmente al organizador del programa. A continuación, el organizador envía a cada laboratorio participante un informe en el que consta, para cada procedimiento de medida sometido a control, las medidas y los coeficientes de variación (imprecisión interdiaria) correspondientes a los resultados enviados el mes anterior y los enviados hasta el momento desde el inicio del periodo de control (datos acumulados), así como la media y la desviación típica correspondiente a los resultados de los laboratorios que utilizan el mismo método (valor consensual por

método) y la media correspondiente a los resultados de todos los laboratorios, prescindiendo del método empleado (valor consensual global) (IUPAC, 1996).

Los resultados obtenidos de un programa de CCEst nunca deben utilizarse como base para disminuir el error sistemático calculando un factor de corrección para aplicarlo a los resultados observados en las muestras de los pacientes, ya que los materiales de control, frecuentemente no se comportan igual que ellos, porque no son de origen humano, porque contienen sustancias conservantes o porque no son equiparables a los procedimientos habituales.

Valores aberrantes: A veces el resultado enviado por un participante es calificado de aberrante por el organizador. En estos casos debe comprobarse que no se ha cometido ninguna equivocación en la transcripción de los resultados, que se han efectuado correctamente los cálculos y que se ha utilizado el material de control adecuado.

Valores incoherentes: En los programas en que se miden dos materiales de control simultáneamente, puede no existir coherencia entre los dos resultados. En este caso, en primer lugar, se debe considerar un posible cruce de los dos materiales de control. Si no existe tal confusión, se debe consultar el diagrama de Youden (Ferré y Fuentes, 1998), que se suele incluir en estos programas, para deducir el tipo de error. En última instancia se debe realizar un control adicional utilizando diferentes materiales de control con valores cercanos a los fisiológicos y a los patológicos para reconocer las influencias de la matriz.

Un punto clave en los programas de CCEst es el establecimiento de los intervalos de aceptación de los resultados. La Organización Internacional para la

Estandarización (ISO) propone cinco métodos para asignar valores de referencia (ISO13528:2015) (Tabla 8).

Tabla 8. Métodos para obtener valores de referencia en un programa de control de calidad externo según la norma ISO13528:2015.

Valor de referencia	Método
1. Formulación	Mezcla de medios
2. Valor de referencia certificado	Materiales certificados CRM
3. Valor de referencia	Comparación con materiales CRM
4. Valor consenso de laboratorios expertos	Valor obtenido por la media de un grupo de laboratorios expertos
5. Valor consenso de los participantes	Valor obtenido por la media de los laboratorios participantes

CRM: Material de referencia certificado.

5.8. Programa de control de calidad externo en el laboratorio de reproducción humana asistida

Munné et al. (1997) encontraron diferencias entre distintos laboratorios en el grado de mosaicismo embrionario al intentar implantar un CCExt basado en el análisis de embriones sobrantes. Estas diferencias podrían estar relacionadas con condiciones de cultivo y/o protocolos de estimulación subóptimos. De esta manera, estudiando la proporción de embriones que presentan mosaicismo en distintos laboratorios, planteaban realizar un CCExt de los sistemas de cultivo. Este tipo de programa nos daría una idea general de la calidad de las clínicas participantes, tanto en lo que a la estimulación se refiere, como a nivel de laboratorio. El problema fundamental de este método son los aspectos ético-legales de destinar embriones viables para este tipo de programas y/o el sesgo que sería el estudiar solo los

embriones sobrantes, ya que no aporta información acerca de los embriones transferidos. Por otro lado, no arroja información acerca de las medidas de corrección a implementar, ya que no se puede conocer el origen del problema.

Los esquemas de CCExt propuestos para el LRHA y más utilizados en diferentes países son: análisis de semen (concentración de espermatozoides, movilidad, vitalidad, morfología espermática y medición de anticuerpos antiespermatozoide), realización de bioensayos para detectar la toxicidad y las condiciones de cultivo subóptimas, y análisis de imágenes para la evaluación de la calidad y destino clínico de los embriones (Elder y Kastrop, 2003; De Jonge et al., 2003; Castilla et al., 2010). En la Tabla 9 se muestran ejemplos de CCExt en el laboratorio de reproducción humana asistida.

Tabla 9. Ejemplos de control de calidad externo para actividades de procedimiento (Fase Técnica) (Castilla y Mantilla et al., 2016).

Área	Proceso	Actividad	Control de calidad externo
Análisis	Análisis de semen	Determinación de la concentración espermática	Semen crioconservado Suspensión bolitas látex Suspensión espermática Videos
		Determinación de la movilidad espermática	Semen crioconservado Videos
		Determinación de la morfología espermática	Portaobjetos teñidos y/o fijados Semen crioconservado Videos y/o fotos

Tabla 9. Ejemplos de control de calidad externo para actividades de procedimiento (Fase Técnica) (Castilla y Mantilla et al., 2016) (Continuación).

Área	Proceso	Actividad	Control de calidad externo
		Determinación de la vitalidad espermática	Portaobjetos teñidos Semen crioconservado Videos
		Realización de un test hiposmótico	Semen crioconservado Videos
		Determinación de anticuerpos anti-espermatozoide directa	Semen crioconservado Videos
		Determinación de anticuerpos anti-espermatozoide indirecta	Sueros
		Recuperación de espermatozoides móviles	Realización de REM con semen congelado
		Clasificación embrionaria	Videos y/o fotos
		Procesamiento de células y tejidos	FIV/ICSI CT IA
Comparación de indicadores con otros centros			

Tabla 9. Ejemplos de control de calidad externo para actividades de procedimiento (Fase Técnica) (Castilla y Mantilla et al., 2016) (Continuación).

Área	Proceso	Actividad	Control de calidad externo
Crioconservación	Crioconservación de gametos (masculino)		Evaluación externa de semen congelado en el laboratorio
	Crioconservación de gametos (femenino)		Evaluación externa de ovocitos vitrificados en el laboratorio
	Crioconservación de embriones		Comparación del % supervivencia embrionaria con otros centros

5.8.1. Control de calidad externo del análisis de semen

La implantación de programas de CCExt de análisis de semen es amplia en muchos países, pues en la última edición del manual de la OMS de análisis de semen se insistía en la participación en estos programas y se facilitaba un listado de programas de este tipo, nacionales e internacionales (OMS, 2010).

En España, tras una experiencia piloto con 18 laboratorios, se instauró en 1999 el primer programa de CCExt auspiciado por ASEBIR, dicho programa estuvo coordinado con el Comité de control de calidad del Grupo de Interés Especial en Andrología de ESHRE y organizado por el laboratorio CEIFER de Granada. Actualmente el material de dicho programa es también distribuido al programa de evaluación externa de la calidad de AEFA y de la Asociación Española de Biopatología Médica y sus resultados procesados en conjunto (Castilla et al., 2005). Este programa es un CCExt tipo I, es decir, consiste en analizar alícuotas de una misma muestra en diferentes laboratorios dos veces al año. El procesamiento de los

resultados se realiza de forma anónima, introduciendo los datos por duplicado con verificación de los mismos en la aplicación informática encargada de su análisis estadístico. Todos los resultados se analizan por técnica y método utilizado, calculándose media, desviación estándar (DS), coeficiente de variación, máximo, mínimo y rango intercuartílico. Se descartan todos los valores que sean $\pm 3DS$ (Álvarez et al., 2005).

La participación en el programa de CCExt para análisis de semen en España ha ido aumentando año tras año. El porcentaje de abandonos anuales ha ido disminuyendo progresivamente.

5.8.2. Control de calidad externo de evaluación embrionaria.

La evaluación de la calidad embrionaria es clave en el laboratorio, ya que afecta a la decisión de cuantos y qué embriones transferir, lo cual está directamente relacionado con la eficacia de un ciclo de FIV y con la probabilidad de embarazo múltiple.

En España entre 2003 y 2007 existía un programa nacional de CCExt de evaluación embrionaria, organizado por CEIFER y auspiciado por ASEBIR, donde los centros participantes debían catalogar embriones en tres categorías (bueno, regular o malo) y decidir el destino clínico de los embriones (transferir, crioconservar o descartar). El análisis de los resultados de estos años permitió demostrar las discrepancias existentes, ya comentadas, en evaluación embrionaria entre centros y un panel de expertos, ya comentadas (Castilla et al., 2010). A partir de 2011, el programa pasa a ser organizado en su totalidad por ASEBIR y en 2012 se incorpora la clasificación estandarizada publicada por ASEBIR en 2008, pasando los

laboratorios a analizar de forma individualizada distintos tipos de variables como número de células del embrión, decisión clínica sobre el mismo, porcentaje de fragmentación, vacuolización, nucleación y otras características que permiten categorizar al embrión mediante la clasificación embrionaria de ASEBIR (Ardoy et al., 2008). Los laboratorios continúan tomando una decisión clínica (transferir, crioconservar o descartar) sobre cada embrión.

Nuestro grupo demostró, que la participación en este tipo de programas aumenta el grado de acuerdo entre laboratorios en la evaluación morfológica embrionaria (Castilla et al., 2010). Debemos recordar en este punto que la manera de participar en los programas de CCExt según la OMS (2010) es mediante un miembro del personal de laboratorio que tenga un aceptable nivel en el programa de CCInt.

Tras cinco años de la incorporación, en 2012, de la catalogación estandarizada de ASEBIR al programa de CCExt de evaluación embrionaria, sería fundamental conocer como ha afectado este cambio a las diferencias existentes entre centros y expertos.

En un programa de CCExt para la evaluación de imágenes de embriones, las alternativas posibles para establecer el valor de referencia, dentro de las normas ISO13528:2015, son los puntos 4 (valor consenso de laboratorios expertos) y 5 (valor consenso de los participantes) expuestos en la Tabla 8. Tanto Arce et al. (2006), que compara los resultados de la evaluación de tres laboratorios de expertos entre ellos y a su vez con los resultados de varios laboratorios locales, como Ruiz de Assin et al., (2009) que compara los resultados entre un grupo de

laboratorios y un grupo de expertos, demostraron que la variabilidad entre expertos era inferior a la observada entre laboratorios. Demostrando la importancia de utilizar valores verdaderos, basados en la opinión de expertos en lugar de basarse los valores de referencia en las opiniones mayoritarias de los centros participantes.

Actualmente, existen programas similares a los descritos para España en Reino Unido, Estados Unidos y Suiza.

5.8.3. Control de calidad externo de evaluación embrionaria con equipos *time-lapse*

Actualmente no existe ningún programa de CCExt para la evaluación embrionaria mediante equipos de *time-lapse*.

Aunque los métodos *time-lapse* proporcionan una gran cantidad de información sobre el desarrollo del embrión, su análisis no está automatizado (Molder et al., 2015). Además, en cuanto a la medición de tiempos en los equipos *time-lapse*, sólo un estudio ha evaluado la variabilidad intra e interobservador (entre tres observadores), reportando poca variabilidad a este respecto (Sundvall et al., 2013). Como ya hemos comentado, la mayoría de los estudios que buscan desarrollar algoritmos de clasificación usando equipos de TL sólo tienen en cuenta el momento del desarrollo embrionario (Meseguer et al., 2011; Conaghan et al., 2013; VerMilyea et al., 2014; Basile et al., 2015; Milewski et al., 2015; Petersen et al., 2016). Sin embargo, estas plataformas también permiten evaluar embriones mediante la evaluación de características morfológicas clásicas, tales como aspecto de los pronúcleos, simetría, fragmentación y/o multinucleación (Gardner et al., 2016). Ahlstrom et al (2016) demostraron que esta estrategia es clínicamente más

útil que un método de selección de embriones en el que solo se analicen los tiempos del desarrollo. Otros autores, utilizando diversas estrategias, han incorporado las características morfológicas clásicas a los algoritmos basados en morfocinética (Conaghan et al., 2013; Diamond et al., 2015; Liu et al., 2016). En la actualidad, poco se sabe sobre el rendimiento de los equipos *time-lapse* en la evaluación de características morfológicas como la fragmentación o multinucleación. Además, no se ha implementado a nivel nacional ni internacional un CCExt, para evaluar la variabilidad entre laboratorios en la determinación de tiempos de morfocinética.

5.8.4. Control de calidad externo para la realización de ensayos de citotoxicidad

El material de laboratorio utilizado en los LRHA no debe tener un efecto negativo en la viabilidad de los gametos y embriones. Con el objetivo de identificar aquellos fungibles tóxicos, diferentes guías recomiendan que el material sea testado previo a su uso (OMS, 2010; ESHRE Guideline Group on Good Practice in IVF Labs et al., 2016). Existen diferentes bioensayos para este fin (Chan et al., 2001; Elder and Kastrop, 2003; Gardner et al., 2005, Kim et al., 2005), siendo el más utilizado por su facilidad y disponibilidad, el test de supervivencia espermática (TSE) (Iemmolo et al., 2005).

La validez del TSE es crucial para la calidad del laboratorio de andrología y embriología, ya que una mala realización de estos bioensayos podría llevar a utilizar materiales y/o medios embriotóxicos, lo que implicaría una disminución de los resultados, o a desechar materiales y/o medios no tóxicos, lo que provocaría costes

extras. En cualquier caso, un buen sistema de bioensayo de materiales y medios llevaría consigo una mejora de resultados con un menor coste. Con objeto de estandarizar los sistemas de bioensayo, diferentes sociedades científicas han implantado programas de CCExt de toxicidad (American Association of Bioanalyst, Fertility Society of Australia, Asociación para el Estudio de la Biología de la Reproducción) (Castilla et al., 2010). En estos programas, los centros participantes reciben diversos fungibles tratados o no con sustancias tóxicas y deben responder si el fungible es tóxico o no. Los laboratorios participantes desconocen totalmente si los materiales han sido tratados o el número de fungibles tratados.

Actualmente, no existe una estandarización de estos programas de CCExt, y los fungibles enviados a los centros para su evaluación son muy diferentes, desde materiales (puntas de pipeta, catéteres de transferencia, pipetas pasteur, placas petri) hasta medios de cultivo. Se han observado diferencias en los resultados de estos programas de CCExt, no pudiéndose descartar que se deba a la gran heterogeneidad de productos utilizados. Hasta la fecha si el tipo de producto utilizado influye en los resultados obtenidos no ha sido evaluado.

En resumen, como hemos visto en esta sección de la introducción, la participación en programas de CCExt es clave para mejorar la calidad del servicio prestado en el LRHA.

6. Interpretación de resultados de estudios sobre reproducibilidad

6.1. Fiabilidad y acuerdo

En el día a día, los clínicos e investigadores deben medir múltiples variables. Entre estas encontramos los estudios de gravedad de una enfermedad, respuestas a tratamientos, confirmación de diagnósticos y evaluación de nuevas herramientas de diagnóstico, siendo esencial, que los resultados de estas medidas sean válidos y reproducibles (de Vet et al., 2006).

Actualmente se han desarrollado diferentes guías, que pretenden estandarizar la forma de analizar los datos derivados de estudios cuyo objetivo es conocer la reproducibilidad entre observadores o entre métodos de medida. Las más relevantes Consenso basado en Estándares para la selección de Instrumentos de Medida en salud (COSMIN) (Mokkink et al., 2010), la Valoración de la Calidad de los Estudios de Fiabilidad (QAREL) (Lucas et al., 2010) o la Guía para Publicar Estudios de Fiabilidad y Acuerdo (GRRAS) (Kottner et al., 2011).

Los principales conceptos que se incluyen dentro del término reproducibilidad son la fiabilidad y el acuerdo. Estos, son conceptos vitales en el desarrollo de clasificaciones y escalas, gestión de la calidad y diseño de ensayos clínicos (Mulsant et al., 2002; Dunn, 2004; Szklo y Nieto, 2014). Los resultados de estudios de fiabilidad y acuerdo proporcionan información sobre la cantidad de error inherente en cualquier diagnóstico, clasificación o medición, donde el tamaño del error de medida determina la validez de los resultados del estudio o de sus

resultados (Dunn, 2004; Shoukri, 2004; Polit y Beck, 2008; Streiner y Norman, 2008; Szklo y Nieto, 2014).

Los términos "fiabilidad" y "acuerdo" a menudo se usan indistintamente. Sin embargo, estos dos conceptos no son lo mismo (Bland y Altman, 1990; Feinstein y Cicchetti, 1990; de Vet et al., 2006; Terwee et al., 2007; Gwet, 2008). La fiabilidad puede definirse como la capacidad de un instrumento o clasificación para diferenciar entre los sujetos/objetos evaluados (Kottner et al., 2011). Por otro lado, el acuerdo es el grado en el cual las puntuaciones o calificaciones coinciden entre los observadores. Ambos conceptos son importantes, porque proporcionan información sobre la calidad de las mediciones.

Aplicados al laboratorio de embriología, los parámetros de fiabilidad valoran como de bien un instrumento o clasificación puede distinguir un embrión de otro. Por ejemplo, basado en alguna característica morfológica, en la clasificación de ASEBIR o según el destino final del embrión (Mokkink et al., 2010). Mientras que los parámetros de acuerdo son utilizados para indicar cuanto coinciden los observadores, al evaluar las mismas características morfológicas embrionarias.

Estos conceptos pueden medirse entre-observadores (evaluadores diferentes, utilizando la misma escala, clasificación, instrumento o procedimiento, evaluando los mismos sujetos u objetos) o intra-observador (el mismo evaluador, con la misma escala, clasificación, instrumento o procedimiento, evalúa los mismos sujetos u objetos en distintos momentos (test-retest)).

Al informar sobre los resultados de fiabilidad y acuerdo, es necesario proporcionar la información suficiente para comprender cómo se diseñó y se llevó a

cabo el estudio y cómo se obtuvieron los resultados. La fiabilidad y el acuerdo no son características fijas de las herramientas de medición, sino, más bien, son el producto de las interacciones entre el contexto de la evaluación, las herramientas y los sujetos/objetos evaluados o medidos.

Las estimaciones de fiabilidad y acuerdo se ven afectadas por diversas fuentes de variabilidad tanto del procedimiento de medida (evaluador/es, tipo de instrumento, momento de medida) como en el enfoque estadístico (modelo estadístico). Por lo tanto, los resultados del estudio sólo son interpretables cuando el ajuste de la medición está suficientemente descrito y el método de cálculo o presentación gráfica se explica completamente.

Como regla general, la mayoría de las guías recomiendan describir los potenciales sesgos que puedan encontrarse en estudios de reproducibilidad (Kottner et al., 2011). En términos epidemiológicos, los sesgos se definen como un error sistemático en el diseño o la realización de un estudio. Pueden aparecer en cualquier etapa del estudio y van desde el método de selección de los participantes (sesgo de selección) a los procedimientos de recopilación de información sobre exposición/enfermedad (sesgo de información). La presencia de sesgos afecta a la validez interna de los estudios y, en consecuencia, hace que las conclusiones obtenidas no sean válidas.

En general, los sesgos en estudios de precisión diagnóstica están bien definidos, y estos mismos pueden ser aplicados a estudios de reproducibilidad. Por ejemplo, sesgo de verificación, "cuando el *gold standard* está influenciado por los resultados de la prueba testada, ya que, por lo general no suelen utilizarse test de

referencia, cuando el resultado es negativo" (Zhou et al.,1998) o sesgo de espectro "cuando los investigadores solo incluyen casos 'claros' o 'definitivos', que no representan todo el espectro de presentación de la enfermedad, o sujetos control sanos que no representan las condiciones en las que un diagnóstico diferencial debería llevarse a cabo".

Es necesario informar de los sesgos potenciales y proporcionar, si es posible, métodos para su corrección.

6.2. Guía para Publicar Estudios de Fiabilidad y Acuerdo (GRRAS)

Ante la ausencia de estándares para realizar estudios de fiabilidad y acuerdo en el campo clínico y con la idea de que estos estándares fueran útiles para investigadores, autores, revisores y editores de revistas, un comité de ocho expertos decidió redactar la Guía para Publicar Estudios de Fiabilidad y Acuerdo (GRRAS) para intentar solucionar esta falta de estandarización. El comité estaba formado por un grupo internacional de expertos avalados por su contribución y publicaciones en el campo de la fiabilidad y el acuerdo, con experiencia en la evaluación y desarrollo instrumental, en estimación de fiabilidad y acuerdo o en revisiones sistemáticas en estudios de fiabilidad.

Además, GRAAS está recogida en la red EQUATOR. Esta red EQUATOR (de las siglas en inglés "mejora de la calidad y la transparencia de las investigaciones de salud") es una iniciativa internacional que busca mejorar la fiabilidad y el valor de la literatura publicada de investigaciones de salud promoviendo la publicación de informes transparentes y precisos y un uso más amplio de guías fiables.

La guía GRAAS contiene 15 puntos que deberían ser incorporados en los estudios de fiabilidad y acuerdo. Los puntos propuestos se recogen en la Tabla 10. Los estudios pueden ser llevados a cabo con el objetivo principal de estimar fiabilidad y acuerdo en sí mismo o pueden ser parte de estudios más grandes sobre precisión diagnóstica, ensayos clínicos o estudios epidemiológicos. Por lo general, en estos últimos, los resultados se informan solamente en un par de frases. Sin embargo, es conveniente que los resultados de fiabilidad y acuerdo se publiquen como parte principal de los estudios.

Tabla 10. Guía para publicar estudios de acuerdo y fiabilidad (Kottner et al., 2011).

TÍTULO Y RESÚMEN	1. Indicar en el título o en el resumen que se analiza la fiabilidad o el acuerdo entre/intra-observador.
INTRODUCCIÓN	2. Nombrar y describir el dispositivo de medición o diagnóstico de interés. 3. Especificar la población de interés 4. Especificar la población de observadores de interés (si corresponde). 5. Describir lo que ya se conoce acerca del acuerdo y fiabilidad y el fundamento para el estudio (si aplica).
MATERIAL Y MÉTODOS	6. Explicar cómo se eligió el tamaño de la muestra. El número de observadores, sujetos/objetos y observaciones. 7. Describir el método de muestreo. 8. Describir el proceso de medición/clasificación (por ejemplo, intervalo de tiempo entre mediciones repetidas, la disponibilidad de información clínica, cegamiento). 9. Si las mediciones/calificaciones se llevaron a cabo independientemente. 10. Describir el análisis estadístico.

Tabla 10. Guía para publicar estudios de acuerdo y fiabilidad (Kottner et al., 2011)
(Continuación).

RESULTADOS	11. El número de observadores y sujetos/objetos evaluados que fueron incluidos y el número de observaciones repetidas que se realizaron. 12. Describir las características de la muestra de evaluadores y de sujetos/objetos evaluados (por ejemplo, la formación, experiencia). 13. Resultados de fiabilidad y de acuerdo, incluidas las medidas de incertidumbre estadística.
DISCUSIÓN	14. Discutir la importancia práctica de los resultados.
MATERIAL AUXILIAR	15. Aportar resultados detallados, si es posible (por ejemplo, en línea).

6.3. Estadísticos recomendados en estudios de fiabilidad y acuerdo

Hay varios enfoques estadísticos que se pueden utilizar en la medición de fiabilidad y acuerdo. Habitualmente, se han desarrollado individualmente para cada tipo de estudio, por tanto, no es posible considerar un único estadístico, como estándar pues dependiendo del estudio, se deberá aplicar un estadístico u otro. Cada estadístico, además, dependerá de las características del tipo de datos a analizar (ordinales, nominales, continuos), del muestreo realizado (al azar, consecutivo, conveniencia) y del tratamiento del error (Dunn et al., 2004, Mokkink et al., 2010). Por tanto, no es posible ser demasiado estrictos sobre cuál es el "mejor" método estadístico, ya que el método a escoger depende de la finalidad y el diseño del estudio.

En la Tabla 11 se enumeran las medidas de fiabilidad y acuerdo según el tipo de variable.

Tabla 11. Métodos estadísticos para el análisis de acuerdo y fiabilidad ente/intra-observadores en estudios de acuerdo (Kottner et al., 2011).

Tipo de variable	Medidas de fiabilidad	Medidas de acuerdo
Nominal	Estadísticos tipo kappa	Proporción de acuerdo
		Proporción de acuerdo específico
Ordinal	Estadísticos tipo kappa ponderados	Proporción de acuerdo
	Matriz de coeficientes tipo kappa	Proporción de acuerdo específico
Continua	Coeficiente de correlación intraclase	Proporción de acuerdo y acuerdo específico
		Error estándar
		Coeficiente de variación
		Gráficos Bland-Altman y límites de acuerdo

6.3.1. Parámetros de fiabilidad para variables cualitativas (nominales, dicotómicas y ordinales)

Los estadísticos tipo kappa (kappa, Gwet, Brennan-Prediger...) son el método más usado para valorar fiabilidad para variables categóricas (Hernaiz, 2015). Proporcionan información útil sobre fiabilidad para datos categóricos

(Kraemer et al., 2002; Dunn, 2004). sin embargo, hay varios tipos de estadísticos kappa, incluyendo kappa de Cohen (para dos observadores), de Weiss (para múltiples observadores), así como sus versiones ponderadas (para variables nominales u ordinales), además, los procedimientos de inferencia también varían dependiendo del estadístico adoptado.

Podemos definir kappa (k) como la proporción de acuerdo más allá del acuerdo esperado por azar (Sim y Wright, 2005). Aunque los índices k han sido ampliamente utilizados, estos, están lejos de ser perfectos.

Por una parte, k está influenciado por la distribución marginal de las clasificaciones y la prevalencia de las variables estudiadas. k asume una distribución equitativa de la prevalencia de las categorías de la variable bajo estudio. Si estas categorías no presentan una prevalencia parecida en la población de estudio los valores de k se distorsionan y pasan a ser menos significativos, este problema es conocido como la paradoja de kappa. (Cicchetti and Feinstein, 1990; Feinstein and Cicchetti, 1990; Hoehler, 2000). Si la prevalencia de una observación positiva es muy alta o muy baja, la probabilidad de acertar por el azar es alta y el valor de k se reduce.

Por otra parte, k , como cualquier estadístico, puede perder alguna información y esta pérdida de información, se acentúa más cuando se convierte para su interpretación de una variable continua a una categórica (Donner et al., 1994), como ocurre con la valoración del índice k propuesta por Landis y Koch (1977) (Tabla 12). Además, los criterios para evaluar k son arbitrarios, aunque por su uso tan frecuente, ha arraigado entre los investigadores los publicados.

Tabla 12. Propuesta de valoración del índice kappa (Landis y Koch, 1977).

	Interpretación del índice Kappa				
	Pobre	Bajo	Moderado	Sustancial	Casi perfecto
Índice kappa	< 0.20	0.21-0.40	0.41-0.60	0.61-0.80	> 0.80

Por tanto, para una correcta interpretación de k es necesario conocer la prevalencia (P) de las categorías de la variable analizada. Se entiende por P , la frecuencia relativa en la que una categoría es elegida por un observador, y por índice de prevalencia (para variables dicotómicas), la diferencia entre las frecuencias en la que una categoría, de una variable dicotómica, es elegida por dos observadores diferentes de forma coincidente, dividida por el número total de observaciones (Sim y Wright, 2005).

Los estadísticos tipo k ponderados, son una extensión de los estadísticos tipo k regulares para variables ordinales, en la que el investigador proporciona pesos arbitrarios para sancionar el grado de desacuerdo entre observadores.

6.3.2. Parámetros de fiabilidad para variables cuantitativas

Para cuantificar la fiabilidad de las mediciones asociadas a las variables cuantitativas continuas, el índice estadístico que se debe utilizar es el coeficiente de correlación intraclase (CCI) (de Vet et al., 2003; Streiner, 2003; de Vet et al., 2006; Gisev et al., 2013). Fleiss y Cohen (1973) demostraron que el CCI es matemáticamente equivalente a los índices k y k ponderado utilizado para variables cualitativas.

Se basa en un modelo de análisis de la varianza (ANOVA) con medidas repetidas (Bland et al., 1990).

El CCI se define como la proporción de variabilidad total debida a la variabilidad de los sujetos/objetos evaluados. Como en el caso de cualquier proporción, los valores del CCI pueden oscilar entre 0 y 1: donde el 0 indica ausencia de concordancia y el 1, la concordancia o fiabilidad absoluta de los resultados obtenidos. Si bien el valor del CCI que define una fiabilidad satisfactoria es arbitrario y depende del uso que de ella se haga, en general, se ha indicado que valores del CCI por debajo del 0,4 representan baja fiabilidad, que valores entre 0,4 y 0,75 representan una fiabilidad entre regular y buena, y que valores por encima de 0,75 representan una fiabilidad excelente (Bland et al., 1990; Hernaez, 2015).

La restricción más importante al uso del CCI es que se trata de una prueba paramétrica y cuando utilizamos pruebas paramétricas debemos considerar las posibles violaciones de las asunciones subyacentes, en nuestro caso normalidad de las distribuciones de las variables, igualdad de varianzas e independencia entre los errores producidos por los observadores.

Otra restricción importante del CCI, al igual que cualquier otro coeficiente de correlación, es que es dependiente de la variabilidad de los valores observados. Si los pacientes varían poco en sus puntuaciones (muestra homogénea), el CCI tiende a ser bajo ya que compara la variabilidad entre pacientes con la varianza total observada, que incluye la varianza de los pacientes, la varianza de los métodos u observadores y el error aleatorio. Si la muestra es heterogénea, el CCI tiende a ser mayor.

6.3.3. Parámetros de acuerdo para variables cualitativas

Se han propuesto como medidas de acuerdo el porcentaje de acuerdo (Fleiss et al., 2003; De Vet et al., 2006), proporciones específicas de acuerdo (Fleiss et al., 2003) y el método gráfico propuesto por Bland y Altman (1990).

Se considera el porcentaje de acuerdo como el número de veces que coinciden varios observadores en una observación entre el número total de observaciones, expresado en forma de tanto por uno.

Para calcular la proporción de acuerdo positivo (el número de veces que coinciden varios observadores en la evaluación de una característica que está presente entre el número total de evaluaciones de dicha característica cuando está presente; P_{pos}) y de acuerdo negativo (el número de veces que coinciden varios observadores en la evaluación de una característica que no está presente entre el número total de evaluaciones de dicha característica cuando no está presente; P_{neg}) (Cicchetti and Feinstein, 1990; Lantz et al., 1996; Hoehler, 2000).

La proporción de acuerdo total junto con las proporciones de acuerdo, dan una estimación del acuerdo global entre evaluadores y para cada categoría (acuerdo positivo o negativo). Las proporciones de acuerdo no se utilizan de forma rutinaria, sin embargo, deberían ser considerados un estándar en la presentación de estadísticas descriptivas básicas, de manera que ayuden a interpretar estadísticos más complejos (Hernaiz, 2015).

6.3.4. Parámetros de acuerdo para variables continuas

El coeficiente de variación (CV) expresa la desviación estándar como un porcentaje de la media. El CV al analizar una sola variable tiene como objetivo describir la dispersión de los valores de la variable de tal manera que, cuanto mayor sea el CV, mayor es la dispersión y por tanto menor acuerdo.

Una de las ventajas del CV es que no tiene unidades lo que puede utilizarse para comparar variables que tuviesen distinta magnitud de medida. La desviación estándar y la media de una variable se expresan en las mismas unidades, por lo que, al calcular el CV, queda sin unidades.

El problema más evidente surge cuando la media de una variable es cero. En este caso, no se puede calcular el CV o en el caso de que la media sea cercana a cero los valores de CV pueden estar distorsionados. El CV puede considerarse como una medida razonable si la variable solo contiene valores positivos (UCLA: Statistical Consulting Group).

En resumen, para obtener conclusiones validas en los estudios de reproducibilidad es necesario aplicar guías estandarizadas para la realización y publicación de estos trabajos. Adaptando los procedimientos a las características específicas del estudio.

Hipótesis y justificación

1. Hipótesis

- La incorporación de una catalogación embrionaria estandarizada a un programa de control de calidad externo de evaluación morfológica embrionaria clásica afectaría a la fiabilidad y acuerdo entre laboratorios y un panel de expertos.
- Por su reciente implantación, los equipos *time-lapse* carecen de controles de calidad externos, lo que podría restar validez a dichos sistemas al no controlarse la variabilidad entre observadores.
- Los diferentes fungibles utilizados en los controles de calidad externo de citotoxicidad podrían influir en la capacidad de los laboratorios de detectar fungibles tóxicos.

2. Justificación

Uno de los retos a los que se enfrentan los laboratorios de reproducción asistida es el mantenimiento de la calidad en todos los procesos llevados a cabo en él. Para ello, se debe evaluar la eficacia de las decisiones y procedimientos realizados diariamente, identificar y corregir problemas, asegurar la certeza y la precisión de los procedimientos, y controlar la competencia del personal del laboratorio. Por este motivo es fundamental analizar y evaluar los procesos y técnicas que se llevan a cabo en el laboratorio de embriología de forma tradicional y aquellos procesos y técnicas nuevas que pretenden incorporarse a dichos laboratorios con objeto de asegurar y garantizar el buen funcionamiento de los mismos.

En los últimos años los dos avances más relevantes en el campo de la clasificación embrionaria han sido la publicación de los criterios estandarizados de catalogación embrionaria de ASEBIR y la aparición de plataformas de morfocinética. Estas últimas se han instalado en nuestros laboratorios sin análisis previos de su reproducibilidad entre centros. Aunque actualmente se ha comprobado una baja variabilidad entre observadores del mismo centro, se desconoce si existen variaciones entre centros a la hora de determinar los tiempos y las características del desarrollo embrionario.

Como se ha comentado anteriormente, es fundamental analizar y evaluar los procesos y técnicas que se llevan a cabo en el laboratorio de embriología. En la actualidad existen a nivel internacional distintos controles de calidad externos para evaluar fungibles gametotóxicos, sin embargo, se desconoce si el rendimiento de un laboratorio en la evaluación de fungibles tóxicos depende del tipo de fungible evaluado.

Objetivos

1. Analizar la reproducibilidad entre laboratorios de embriología y un panel de expertos en la evaluación morfológica embrionaria clásica.
2. Analizar el impacto de los equipos *time-lapse* en la variabilidad entre laboratorios de embriología en la evaluación morfológica embrionaria.
3. Valorar la influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad en el laboratorio de embriología.

Material y métodos

1. Acuerdo y fiabilidad entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica

Para este objetivo se comparan las respuestas dadas, por los laboratorios participantes y un panel de expertos, durante cinco años, en el programa de CCEExt para el laboratorio de reproducción, organizado por ASEBIR con la colaboración de CEIFER utilizando una catalogación embrionaria estandarizada.

1.1. Participación

El número de laboratorios participantes en el programa del CCEExt de evaluación morfológica embrionaria, durante el periodo 2012-2016, ha sido de 42, 39, 39, 55 y 67 laboratorios, en los años 2012, 2013, 2014 2015 y 2016, respectivamente.

En todos los casos la participación fue voluntaria e independiente de otros programas de CCEExt del laboratorio de reproducción organizados o auspiciados por ASEBIR (citotoxicidad, análisis de semen).

1.2. Videos

Los videos se realizaron utilizando una cámara de video conectada a un microscopio de contraste de fases [Microscopio invertido- Cámara de video (Nikon, Tokio-Japón)]. La óptica y el aumento utilizado para las grabaciones fueron Hoffman y 400x, respectivamente. Las grabaciones se llevaron a cabo según la rutina diaria de evaluación de la calidad embrionaria y son seleccionadas a criterio de los organizadores del CCEExt para el laboratorio de reproducción. Cada video tiene una duración de entre 15 y 35 segundos. En ellos se realiza un barrido a lo largo de

todos los planos focales de forma automática de manera que pueda observarse la totalidad del embrión.

Los videos de los embriones utilizados para el programa de CCEExt son obtenidos de los programas de FIV/ICSI del Centro MasVida Reproducción (Sevilla). Las pacientes habían autorizado previamente, la utilización de las imágenes de sus embriones de manera anonimizada con fines de divulgación científica.

1.3. Visualización de las imágenes

Los laboratorios acceden a las imágenes a través de una plataforma online. En ella las imágenes están agrupados en cinco bloques (punciones):

- Punción 1, 2 y 3: con 15 vídeos cada punción (5 videos de cigotos, 5 videos de embriones en día 2 y 5 videos de embriones en día 3).
- Punción 4 y 5: con 10 vídeos cada punción (5 videos de embriones en Día 4 y 5 videos de embriones en día 5).

1.4. Boletín de resultados

Los laboratorios participantes y el panel de expertos debían cumplimentar un boletín de resultados estructurado de la siguiente manera:

Para embriones en día 2 y día 3:

- Número de células.
- Asimetría de las blastómeras (en estadio de 2, 4 u 8 células).
- Fragmentación celular: Porcentaje de fragmentos (0 a 100% en intervalos de 5%).

Esta variable se agrupó en dos categorías ($\leq 10\%$ o $> 10\%$) para su análisis estadístico con el objeto de tener poblaciones de embriones con una representación balanceada de la característica analizada.

- Presencia de vacuolas (presencia o ausencia).
- Multinucleación (presencia o ausencia).

Para embriones en día 5:

- Masa celular interna según catalogación ASEBIR (A, B, C o D)
- Trofoectodermo según catalogación ASEBIR (A, B, C o D)

Así como la catalogación embrionaria según criterios ASEBIR (A, B, C o D) y destino clínico del embrión tras su cultivo (descartar, crioconservar o transferir) para embriones en día 2, 3 y 5.

Para las punciones 1, 2 y 3 los participantes deben elegir solo dos embriones para transferir y decidir para el resto de los embriones cualquiera de los otros destinos posibles (descartar o crioconservar).

Para las punciones 4 y 5 los participantes deben elegir solo un blastocisto para transferir y decidir para el resto, cualquiera de los otros destinos posibles (descartar o crioconservar).

Para estas dos últimas punciones, los laboratorios fueron informados de que los embriones habían sido catalogados como embriones tipo A en día 3. Esta información se dio para conocer la fiabilidad y el acuerdo en la evaluación

morfológica de blastocistos y en el destino de los embriones evitando sesgos que podrían producirse debido a la evaluación del desarrollo previo del embrión.

1.5. Evaluación de los videos

Se considera para cada variable analizada en la evaluación realizada por los laboratorios, la respuesta dada por la mayoría de los laboratorios participantes.

En ausencia de un “*gold standard*”, la categoría embrionaria, la evaluación de las características morfológicas y el destino de los embriones fue definido por consenso entre un panel de expertos. Este panel está constituido por siete miembros, escogidos debido a su calidad de expertos en el ámbito de la embriología humana en España. Estos expertos debían ser embriólogos séniores de ASEBIR y ESHRE, tener al menos 10 años de experiencia en el campo de la embriología clínica y ser miembros de los grupos de interés de Calidad o Embriología de ASEBIR. Estos siete embriólogos se reúnen de forma presencial y anual para analizar las mismas imágenes que se les mostraron a los laboratorios.

En los cinco años analizados, un total de 125 embriones fueron evaluados (25 por año). Este resultó en un total de 200 evaluaciones morfológicas embrionarias (40 por año; 15 en día 2, 15 en día 3 y 10 en día 5 embriones) y 200 decisiones sobre el destino de los embriones (40 por año; 15 en día 2, 15 en día 3 y 10 en día 5 embriones) (Tabla 13).

Tabla 13. Actividad anual del programa de control de calidad externo, días en los que los centros deben realizar la evaluación morfológica embrionaria y determinar el destino del embrión y el número de evaluaciones hechas.

	Evaluación de las imágenes					Evaluaciones
	Cigoto	Día 2	Día 3	Día 4	Día 5	
Punción 1 (5 embriones)						
EMEC	No	Sí	Sí	—	—	10
DE	No	Sí	Sí	—	—	10
Punción 2 (5 embriones)						
EMEC	No	Sí	Sí	—	—	10
DE	No	Sí	Sí	—	—	10
Punción 3 (5 embriones)						
EMEC	No	Sí	Sí	—	—	10
DE	No	Sí	Sí	—	—	10
Punción 4 (5 embriones)						
EMEC	—	—	—	No	Sí	5
DE	—	—	—	No	Sí	5
Punción 5 (5 embriones)						
EMEC	—	—	—	No	Sí	5
DE	—	—	—	No	Sí	5

EMEC, evaluación morfológica embrionaria clásica; DE, destino del embrión; Evaluaciones (por año).

Ni las imágenes de los cigotos ni de los embriones en día 4 fueron analizadas, solo se incluyeron en el programa de CCEExt para dar información en la toma de decisiones sobre el destino de los embriones.

2. Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante *time-lapse*

Los datos utilizados para este objetivo fueron obtenidos de dos fuentes:

- Ensayo piloto del programa español de CCExt para el laboratorio de reproducción utilizando equipos *time-lapse*, organizado por ASEBIR en el año 2013.
- Los datos obtenidos de los laboratorios participantes en el programa de CCExt de EMEC durante los años 2013-2014 organizado por ASEBIR con la colaboración de CEIFER.

2.1. Participación

El número de laboratorios participantes en el ensayo piloto de un programa del CCExt para el laboratorio de reproducción en la evaluación morfológica embrionaria usando equipos *time-lapse*, ha sido de 24 laboratorios (8 laboratorios han participado con el equipo EmbryoScope™ (Unisense FertiliTech, Denmark), 15 con el equipo Primo Vision™ Version 4.3.1.5 (Vitrolife, Sweden) y uno participó con ambos equipos).

Como ya se ha comentado, en todos los casos la participación fue voluntaria e independiente de otros programas de CCExt del laboratorio de reproducción organizados o auspiciados por ASEBIR (citotoxicidad, análisis de semen).

2.2. Videos

Los videos de los embriones utilizados para el primer ensayo piloto de un programa de CCExt para equipos *time-lapse* fueron obtenidos con los equipos EmbryoScope™ o Primo Vision™.

Estos videos, se generan a partir de fotografías realizadas al embrión por los equipos *time-lapse*. Estas fotografías capturan distintos planos focales del embrión, cada 5-20 minutos.

Para la plataforma Embryoscope™ los videos fueron obtenidos con ayuda de un software que acompaña al sistema. Este software genera un video *time-lapse* a partir de imágenes estáticas de alta definición que se obtienen cada 10 minutos, las imágenes son tomadas a lo largo de siete planos focales. Estos embriones fueron cultivados a una concentración de oxígeno del 5%.

Los videos *time-lapse* para la plataforma PrimoVision, se generaron a partir de imágenes que constan de siete planos focales de 200 µm y son capturadas cada 5 minutos durante las primeras 24 horas y cada 20 minutos hasta el día de la transferencia embrionaria. Estos embriones fueron cultivados a una concentración de oxígeno del 21%.

Los videos de los embriones utilizados para el primer ensayo piloto de un programa de CCExt para equipos *time-lapse* fueron obtenidos de los programas de ICSI de GineFIV (Madrid), Clínica Tambre (Madrid) y Unidad de Reproducción del Hospital Universitario Virgen de las Nieves (Granada).

2.3. Esquema organizativo del ensayo piloto de control de calidad externo de evaluación morfológica embrionaria utilizando equipos *time-lapse*

En 2014 se inició un programa piloto con el objetivo de comprobar la viabilidad de la instauración de un programa de CCExt para el laboratorio de reproducción utilizando equipos *time-lapse*, que se realizó de una manera totalmente gratuita. Cada centro recibió los videos de los ovocitos microinyectados de tres punciones, en un dispositivo USB, incluyendo cada punción, videos del desarrollo de entre 9-14 ovocitos hasta la fase de blastocisto. Los usuarios de Embryoscope analizaron un total de 31 embriones y los usuarios de PrimoVision analizaron un total de 35 embriones.

La distribución del pendrive a los diferentes laboratorios participantes se realizó a través de mensajería (empresa de transportes MRW) en modalidad de envío ordinario.

Para la distribución, se utilizaron sobres acolchados con el pendrive debidamente identificado. Asimismo, se incluían en el sobre las características y especificaciones del ensayo piloto del programa de CCExt utilizando equipos *time-lapse*.

2.4. Boletín de resultados

Los laboratorios participantes debían cumplimentar un Boletín de resultados estructurado de la siguiente manera:

Una primera parte en la que se recogen características morfológicas cualitativas:

- Numero de pronúcleos.
- Fragmentación celular: Porcentaje de fragmentos (0%, <10%, 11-25%, 26-35% >35%).
- Presencia de vacuolas: inclusiones citoplasmáticas rodeadas de membrana y rellenas de un fluido virtualmente idéntico al del espacio perivitelino (ninguna, escasas o abundantes).
- Asimetría de las blastómeras (en estadio de 2, 4 u 8 células),
- Compactación: formación de uniones intercelulares y el inicio de la polarización embrionaria (Si o No).
- Catalogación embrionaria: Los laboratorios debían clasificar los embriones siguiendo los criterios de ASEBIR.
- Destino clínico del embrión tras el cultivo de los embriones: descartar, crioconservar o transferir.

Una segunda parte en la que se recogen los tiempos en los que se producen los siguientes eventos del desarrollo embrionario (características cuantitativas):

- La extrusión del segundo corpúsculo polar (tPB2).
- La aparición (tPNa) y desaparición (tPNf) de los pronúcleos.
- Cada una de las divisiones celulares (tSt2, t2, t3, t4, t5, t6, t7, t8, t9).
- Tiempo en el que alcanza la fase de mórula (tM).

- Tiempo en el que aparecen los primeros signos de blastulación (tSB).
- Tiempo en el que alcanza la fase de blastocisto (tB).
- Tiempo en el que comienza la expansión (tE).
- Tiempo en el que inicia la eclosión (tHN).
- Tiempo en el que ha eclosionado (tH)

3. Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad

Todos los datos utilizados para el análisis fueron obtenidos del programa de CCExt de citotoxicidad (CT) para el laboratorio de reproducción, durante el periodo 2013-2016, organizado ASEBIR con la colaboración de CEIFER.

Se comparan las respuestas dadas por los laboratorios participantes y el tratamiento (tóxico o no tóxico) dado a los productos por parte de la organización del CCExt de CT.

3.1. Participación

El número de laboratorios participantes en el programa del CCExt para el laboratorio de reproducción en la evaluación de fungibles citotóxicos, durante el periodo 2013-2016, ha sido una media de 48 laboratorios. 37, 39, 54 y 64 laboratorios han evaluado puntas de pipeta y 35, 40, 53 y 63 han evaluado medios de cultivo embrionario en los años 2013, 2014, 2015 y 2016, respectivamente.

Como ya se ha comentado, en todos los casos la participación fue voluntaria e independiente de otros programas de CCExt del laboratorio de reproducción organizados o auspiciados por ASEBIR (citotoxicidad, análisis de semen).

3.2. Esquema organizativo del control de calidad externo de citotoxicidad

En 2003 se puso en marcha, por primera vez, un CCExt para la evaluación de material citotóxico. En sus inicios, los laboratorios evaluaban exclusivamente materiales (Castilla et al., 2010). A partir de 2013 cada envío consistió en la evaluación de medios de cultivo y material (puntas de pipeta).

Los datos de este programa de CCExt de CT fueron comparados con los obtenidos en este mismo programa durante los años 2003-07 (Castilla et al., 2010) y con los obtenidos en el programa de CCExt de CT de la Asociación Americana de Bioanalistas (AAB) (AAB, 2017). En este último caso se analizan los resultados obtenidos durante los años 2013 a 2016.

3.3. Productos y reactivos para el chequeo por bioensayos

3.3.1. Productos

- Puntas de pipeta amarillas (Daslab, España)
- Medio de cultivo M2 con HEPES (Sigma-Aldrich, E.E.U.U.)

3.3.2. Reactivos

Como reactivo para hacer citotóxicos a los productos que debían ser analizados en el CCExt de CT, se utilizó clorhexidina. La clorhexidina es un líquido

desinfectante de la familia de las biguanidas. Al contacto con material biológico, causa la precipitación de ácidos nucleicos, desnaturalización de proteínas y lisis celular (Elder et al., 2004).

3.3.3. Envios

La distribución del material a los diferentes laboratorios participantes se realizó a través de mensajería (empresa de transportes MRW) en modalidad de envío ordinario.

Para la distribución del material se utilizaron sobres acolchados con el material fungible debidamente etiquetado. Asimismo, se incluían en el sobre las características y especificaciones del programa de CCEExt.

Se envió anualmente a los laboratorios participantes 3 muestras de medios de cultivo embrionario y 3 puntas de pipeta. Los resultados de una de las puntas de pipeta no se incluyen en este estudio ya que su tratamiento no fue consistente a lo largo de los años analizados.

El envío de las puntas de pipeta se realizó a temperatura ambiente mientras que el de medios de cultivo se realizó en condiciones refrigeradas. Los envíos se realizaron siempre en lunes para evitar demoras en la entrega del material. Los laboratorios fueron instruidos para conservar los medios refrigerados hasta su evaluación, la cual debía realizarse lo antes posible.

3.4. Tratamiento de los productos a analizar

Algunos de esos materiales fueron tratados por la organización de tal manera que resultasen tóxicos. Los medios se hicieron tóxicos mediante la adición de clorhexidina 0,01% y las puntas se incubaron con clorhexidina 2% durante 24h.

Los laboratorios participantes desconocían totalmente cuantos de los productos habían sido o no tratados.

Los medios fueron manipulados en condiciones de esterilidad y las puntas de pipeta fueron esterilizadas en autoclave. Todos los materiales habían sido testados previamente por los organizadores del programa, los cuales comprobaron mediante test de supervivencia espermática (TSE) la toxicidad de los productos enviados, observando reducciones superiores al 50% de la movilidad inicial a las 24 h.

3.5. Boletín de resultados

Los resultados de cada centro fueron remitidos a través de una web desarrollada con ese fin.

Los laboratorios participantes debían decir si los productos testados eran tóxicos o no tóxicos para su uso en el laboratorio de reproducción.

Todos los laboratorios utilizaron, para valorar la toxicidad de los fungibles, el TSE.

4. Métodos estadísticos

4.1. Fiabilidad y acuerdo entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica

Se analizó fiabilidad y acuerdo entre laboratorios y el panel de expertos calculando diferentes estadísticos y sus respectivos intervalos de confianza al 95%, siguiendo las recomendaciones de la GRRAS.

4.1.1. Fiabilidad

La kappa de Cohen se calculó como medida de fiabilidad entre laboratorios y el panel de expertos. Así como prevalencia (P) (frecuencia con la que una categoría en particular fue indicada por el panel de expertos) y el índice de prevalencia (IP) (la diferencia entre las frecuencias de acuerdo entre laboratorios y un panel de expertos en la presencia o ausencia de una característica dada; solo usada para variables dicotómicas) (Sim y Wright, 2005) para una correcta interpretación de kappa.

Para evitar la paradoja de kappa se calculó el coeficiente Gwet para dos observadores (Wongpakaran et al., 2013).

Para optimizar la interpretación de Gwet y kappa fue necesario tener en cuenta los conceptos de resultados positivos (aquellas resultados en las que al menos uno de los observadores detecta la presencia de la característica estudiada) y de resultados negativos (aquellas en las que al menos uno de los observadores no detectó la presencia de la característica), y calcular la proporción de acuerdo positivo (el número de resultados positivos coincidentes entre los laboratorios y el

panel de expertos, dividido por todos los resultados positivos; Ppos) y proporción de acuerdo negativo (el número de resultados negativos coincidentes entre los laboratorios y el panel de expertos, dividido por todos los resultados negativos; Pneg) (Cicchetti and Feinstein, 1990; Lantz et al., 1996; Hoehler, 2000).

Para la interpretación el índice Gwet y los valores de kappa, nos basamos en la propuesta de valoración del índice kappa de Landis y Koch (1977) que aparece reflejada en la Tabla 12.

4.1.2. Acuerdo

Para las variables dicotómicas categóricas y para las variables no ordinales policotómicas categóricas se calculó la proporción cruda de acuerdo. Para las variables ordinales policotómicas categóricas se calculó la proporción de acuerdo ponderado.

Para los cálculos anteriores, se utilizó el programa estadístico AgreeStat2015 [Advanced Analytics, LLC (USA)]. Se utilizó una ponderación cuadrática cuando fue necesaria aplicarla en algún parámetro. Para una correcta interpretación del acuerdo ponderado se utilizaron los siguientes intervalos: 0.70-0.79 “pobre”; 0.80-0.89 “moderado”; 0.90-0.98 “fuerte”; 0.99-1.00 “casi perfecto” (Tworek et al., 2013).

4.1.3. Número de observaciones

El número de observaciones necesario para obtener valores de fiabilidad se calculó según la expresión $2n^2$, donde n es el número de categorías para cada característica evaluada. En este caso, el número más alto de categorías (n=4) nos indica la necesidad de tener al menos, $2*4^2=32$ embriones (Paternot et al., 2011).

Un total de 125 embriones fueron incluidos en este trabajo para valorar el acuerdo y la fiabilidad entre laboratorios y un panel de expertos.

4.2. Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante *time-lapse*

Para determinar la variabilidad entre laboratorios en la evaluación de variables cualitativas se utilizó el programa estadístico AgreeStat2015 [Advanced Analytics, LLC (USA)] para calcular Gwet para múltiples observadores, así como sus respectivos intervalos de confianza al 95%.

Para la interpretación del índice Gwet nos basamos en la propuesta de valoración de índices tipo kappa de Landis y Koch (1977) descritos anteriormente en la Tabla 12.

Se consideró un alto grado de acuerdo cuando sobre un embrión o cigoto más del 75% de los laboratorios coincidían en su clasificación o decisión clínica.

Para la comparación de las diferentes variables cualitativas analizadas se utilizó el test de χ^2 con una significación del 5%. En caso de no cumplirse las condiciones de validez de dicho test, se utilizó el test de Fisher.

Los datos obtenidos al analizar las variables cuantitativas se analizaron con el paquete estadístico SPSS 19.0. Para cada tiempo analizado, se calculó el coeficiente de variación (CV) y el coeficiente de correlación intraclase (CCI) junto con sus correspondientes intervalos de confianza al 95%.

De todas las variables cuantitativas se presenta su media y desviación estándar. Los tiempos medios de cada evento para los embriones analizados con cada equipo *time-lapse* se compararon con el test de Student.

4.3. Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad

Para analizar las respuestas dadas por los laboratorios durante los cuatro años estudiados del CCExt de CT, se calculó sensibilidad (S), especificidad (E), valor predictivo positivo (VPP), valor predictivo negativo (VPN) y el acuerdo total de los bioensayos, así como su intervalo de confianza al 95%.

Se define, S, como la proporción de materiales tóxicos que es detectado por el bioensayo; E, como la proporción de materiales no tóxicos detectados por el bioensayo; VPP, como la proporción de bioensayos que detectan toxicidad y este resultado corresponde con un material tóxico, VPN, como la proporción de bioensayos que no detectan toxicidad y este resultado corresponde con un material no tóxico y porcentaje de acuerdo total como la proporción de bioensayos que detectan correctamente el material, ya sea tóxico o no tóxico.

Se analizó el solapamiento de los intervalos de confianza al 95, para conocer la significancia existente entre los estadísticos obtenidos al estudiar los datos del CCExt de CT de ASEBIR durante los años 2013-16, los obtenidos en este mismo programa durante los años 2003-07 (Castilla et al., 2010) y los obtenidos en el programa de CCExt de CT de la Asociación Americana de Bioanalistas (AAB) (AAB, 2017).

Se utilizó una representación del espacio ROC para interpretar los datos del TSE. El espacio ROC es una representación gráfica de la sensibilidad frente a uno menos la especificidad. De igual manera se representaron gráficamente los resultados obtenidos con los dos productos en términos de valor predictivo positivo y valor predictivo negativo.

Para determinar si existía relación entre el porcentaje de acierto al analizar medios de cultivo y el obtenido al analizar puntas de pipeta, se calculó el coeficiente de Correlación de Pearson.

Resultados

1. Acuerdo y fiabilidad entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica

1.1. Catalogación embrionaria

El acuerdo global entre laboratorios y expertos en catalogación embrionaria según criterios de ASEBIR fue sustancial (k : 0.71 (0.64-0.81); G_{wet} : 0.76 (0.70-0.84)). Mediante el cálculo de k no se observa ninguna diferencia en la fiabilidad en la evaluación embrionaria según el día del desarrollo (día 2: 0.66 (0.49-0.83); día 3: 0.75 (0.61-0.89) y día 5: 0.68 (0.53-0.82)) Sin embargo, las elevadas diferencias en la prevalencia de las distintas categorías de calidad embrionaria analizadas en día 5 (14% de embriones tipo A y un 44% de embriones tipo B, según los expertos) hace necesario el cálculo de G_{wet} para evitar la comentada “paradoja de kappa”. Mediante al cálculo de G_{wet} si se observa una ligera tendencia a un mayor grado de fiabilidad entre laboratorios y expertos al analizar embriones con más días de desarrollo (día 2: 0.72 (0.57- 0.87); día 3: 0.79 (0.67-0.91) and día 5: 0.86 (0.78-0.93)). Este menor grado de fiabilidad en día 2 y día 3 se debe a que el porcentaje de acuerdo ponderado entre laboratorios y expertos fue menor (pobre y moderado, respectivamente) en embriones de categoría D que en el resto de categorías. Sin embargo, en las otras categorías este porcentaje de acuerdo fue superior a 0.9 (fuerte) en cualquier día del desarrollo (Tabla 14).

Tabla 14. Acuerdo y fiabilidad entre laboratorios y el panel de expertos en evaluación morfológica embrionaria.

Evaluación morfológica embrionaria	n	P	Acuerdo ponderado	Kappa	Gwet
Día 2					
A	28	37.3	0.92 (0.86-0.98)		
B	12	16.0	0.98 (0.96-1.00)		
C	20	26.7	0.96 (0.91-1.00)		
D	15	20.0	0.78 (0.60-0.96)		
Total	75		0.91 (0.87-0.96)	0.66 (0.49-0.83)	0.72 (0.57-0.87)
Día 3					
A	14	18.7	0.96 (0.89-1.00)		
B	11	14.7	0.96 (0.89-1.00)		
C	23	30.7	0.96 (0.92-1.00)		
D	27	36.0	0.89 (0.80-0.98)		
Total	75		0.93 (0.90-0.97)	0.75 (0.61-0.89)	0.79 (0.67-0.91)

P, prevalencia; (A, B, C, D; categorías de la catalogación ASEBIR donde A tiene el mejor pronóstico y D el peor; (95% intervalo de confianza).

Tabla 14. Acuerdo y fiabilidad entre laboratorios y el panel de expertos en evaluación morfológica embrionaria (continuación).

Evaluación morfológica embrionaria	n	P	Acuerdo ponderado	Kappa	Gwet
Día 5					
A	7	14.0	0.92 (0.87-0.97)		
B	22	44.0	0.94 (0.90-0.99)		
C	13	26.0	0.98 (0.96-1.00)		
D	8	16.0	0.93 (0.89-0.98)		
Total	50		0.95 (0.93-0.97)	0.68 (0.53-0.82)	0.85 (0.78-0.93)
Global					
A	49	24.5	0.93 (0.89-0.97)		
B	45	22.5	0.96 (0.93-0.98)		
C	56	28.0	0.96 (0.94-0.99)		
D	50	25.0	0.86 (0.80-0.93)		
Total	200		0.93 (0.91-0.95)	0.71 (0.64-0.81)	0.76 (0.70-0.84)

P, prevalencia; (A, B, C, D; categorías de la catalogación ASEBIR donde A tiene el mejor pronóstico y D el peor; (95% intervalo de confianza).

1.2. Características morfológicas embrionarias

Para analizar las discrepancias observadas en catalogación embrionaria en día 2 y día 3 entre laboratorios y expertos, determinamos el grado de acuerdo y la fiabilidad en la evaluación de las diferentes variables morfológicas que llevan a la catalogación embrionaria según ASEBIR. Las grandes diferencias en la prevalencia de algunas características morfológicas impiden una correcta interpretación de k . Por lo que se utilizó el índice Gwet. Los índices Gwet obtenidos para las variables morfológicas fragmentación y asimetría, fueron casi perfectos, así como altos porcentajes de acuerdo. En fragmentación y asimetría la P_{pos} y P_{neg} fueron similares, siempre superiores a 0.88 (Tablas 15 y 16).

Tabla 15. Acuerdo y fiabilidad entre laboratorios y el panel de expertos en la evaluación del porcentaje de fragmentación.

Fragmentación	n	P	IP	Acuerdo crudo	Pneg	Ppos	Kappa	Gwet
Día 2								
≤10%	59	78.7		0.95 (0.89-1.00)				
>10%	16	21.3		0.88 (0.70-1.00)				
Total	75		0.56 (0.45-0.67)	0.93 (0.88-0.99)	0.96 (0.91-1.00)	0.85 (0.77-0.93)	0.81 (0.64-0.97)	0.90 (0.81-0.99)
Día 3								
≤10%	59	78.7		0.97 (0.92-1.00)				
>10%	16	21.3		0.94 (0.88-1.00)				
Total	75		0.56 (0.45-0.67)	0.96 (0.92-1.00)	0.97 (0.94-1.00)	0.91 (0.84-0.97)	0.88 (0.75-1.00)	0.94 (0.87-1.00)
Global								
≤10%	118	78.7		0.96 (0.92-0.99)				
>10%	32	21.3		0.91 (0.80-1.00)				
Total	150		0.56 (0.48-0.64)	0.95 (0.91-0.98)	0.97 (0.94-0.99)	0.88 (0.83-0.93)	0.84 (0.74-0.95)	0.92 (0.86-0.98)

P, prevalencia; IP, índice de prevalencia; Pneg, proporción de acuerdo negativo; Ppos, proporción de acuerdo positivo; (95% intervalo de confianza)

Tabla 16. Acuerdo y fiabilidad entre laboratorios y el panel de expertos en la evaluación de la asimetría.

Asimetría	n	P	IP	Acuerdo crudo	Pneg	Ppos	Kappa	Gwet
Día 2								
Igual	32	71.1		1.00 (1.00-1.00)				
Desigual	13	28.9		1.00 (1.00-1.00)				
Total	45		0.42 (0.28-0.57)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
Día 3								
Igual	26	57.8		0.92 (0.82-1.00)				
Desigual	19	42.2		0.95 (0.84-1.00)				
Total	45		0.13 (0.03-0.23)	0.93 (0.86-1.00)	0.94 (0.87-1.00)	0.92 (0.85-1.00)	0.86 (0.71-1.00)	0.87 (0.72-1.00)
Global								
Igual	58	64.4		0.97 (0.92-1.00)				
Desigual	32	35.6		0.97 (0.91-1.00)				
Total	90		0.28 (0.19-0.37)	0.97 (0.93-1.00)	0.97 (0.94-1.00)	0.95 (0.91-1.00)	0.93 (0.85-1.00)	0.94 (0.87-1.00)

P, prevalencia; IP, índice de prevalencia; Pneg, proporción de acuerdo negativo; Ppos, proporción de acuerdo positivo; (95% intervalo de confianza).

En la evaluación de multinucleación, se observó un elevado porcentaje de acuerdo entre laboratorios y expertos cuando esta característica no estaba presente (0.99) pero un bajo porcentaje de acuerdo cuando la multinucleación estaba presente (0.41). Al comparar la probabilidad del positivo (P_{pos}) y probabilidad del negativo (P_{neg}) observamos que son muy diferentes al analizar multinucleación (0.56 (0.48-0.64) vs 0.96 (0.93-0.99)) (Tabla 17).

Tabla 17. Acuerdo y fiabilidad entre laboratorios y el panel de expertos en la evaluación de multinucleación.

Multinucleacion	n	P	IP	Acuerdo crudo	Pneg	Ppos	Kappa	Gwet
Día 2								
Ausencia	64	85.3		1.00 (1.00-1.00)				
Presencia	11	14.7		0.45 (0.12-0.79)				
Total	75		0.79 (0.69-0.88)	0.92 (0.86-0.98)	0.96 (0.91-1.00)	0.63 (0.52-0.73)	0.59 (0.29-0.88)	0.90 (0.82-0.98)
Día 3								
Ausencia	69	92		0.99 (0.96-1.00)				
Presencia	6	8		0.33 (0.00-0.83)				
Total	75		0.88 (0.81-0.95)	0.93 (0.88-0.99)	0.96 (0.92-1.00)	0.44 (0.33-0.56)	0.41 (0.00-0.83)	0.92 (0.86-0.99)
Global								
Ausencia	133	88.7		0.99 (0.98-1.00)				
Presencia	17	11.3		0.41 (0.16-0.67)				
Total	150		0.83 (0.77-0.89)	0.93 (0.89-0.97)	0.96 (0.93-0.99)	0.56 (0.48-0.64)	0.53 (0.28-0.77)	0.91 (0.86-0.97)

P, prevalencia; IP, índice de prevalencia; Pneg, proporción de acuerdo negativo; Ppos, proporción de acuerdo positivo; (95% intervalo de confianza).

En la evaluación de vacuolas se observó un elevado porcentaje de acuerdo entre laboratorios y expertos cuando esta característica no estaba presente (1.00) pero un bajo porcentaje de acuerdo cuando el embrión estaba vacuolado (0.27). Al comparar la probabilidad del positivo (P_{pos}) y probabilidad del negativo (P_{neg}) observamos que son muy diferentes al evaluar presencia de vacuolas (0.43 (0.35-0.51) vs 0.97 (0.94-1.00)) (Tabla 18).

Tabla 18. Acuerdo y fiabilidad entre laboratorios y el panel de expertos en la evaluación de vacuolas.

Vacuolas	n	P	IP	Acuerdo crudo	Pneg	Ppos	Kappa	Gwet
Día 2								
Ausencia	73	97.3		1.00 (1.00-1.00)				
Presencia	2	2.7		0.50 (0.00-1.00)				
Total	75		0.96 (0.92-1.00)	0.99 (0.96-1.00)	0.99 (0.97-1.00)	0.67 (0.56-0.78)	0.66 (0.03-1.00)	0.99 (0.96-1.00)
Día 3								
Ausencia	66	88.0		1.00 (1.00-1.00)				
Presencia	9	12.0		0.22 (0.00-0.54)				
Total	75		0.85 (0.77-0.93)	0.91 (0.84-0.97)	0.95 (0.90-1.00)	0.36 (0.25-0.47)	0.33 (0.00-0.69)	0.89 (0.81-0.98)
Global								
Ausencia	139	92.7		1.00 (1.00-1.00)				
Presencia	11	7.3		0.27 (0.00-0.57)				
Total	150		0.91 (0.86-0.95)	0.95 (0.91-0.98)	0.97 (0.94-1.00)	0.43 (0.35-0.51)	0.41 (0.09-0.73)	0.94 (0.90-0.98)

P, prevalencia; IP, índice de prevalencia; Pneg, proporción de acuerdo negativo; Ppos, proporción de acuerdo positivo; (95% intervalo de confianza).

En la evaluación del número de células, la fiabilidad fue casi perfecta, tanto con el estadístico k (0.99 (0.98-1.00)) como con Gwet (1.00 (1.00-1.00)). Sin embargo, en los embriones con un número de células ≥ 10 según los expertos, el porcentaje de acuerdo fue moderado (0.82 (0.72-0.93)). El porcentaje de acuerdo observado al analizar embriones con un número de células ≤ 10 fue alto, siempre superior al 0.90 (Tabla 19).

Tabla 19. Acuerdo y fiabilidad entre laboratorios y el panel de expertos en la evaluación del número de células.

Nº de células	n	P	Acuerdo ponderado	Kappa	Gwet
2	9	6.0	1.00 (1.00-1.00)		
3	4	2.7	1.00 (1.00-1.00)		
4	53	35.3	1.00 (1.00-1.00)		
5	8	5.3	1.00 (1.00-1.00)		
6	17	11.3	0.93 (0.80-1.00)		
7	11	7.3	0.98 (0.93-1.00)		
8	34	22.7	0.99 (0.98-1.00)		
9	7	4.7	0.93 (0.82-1.00)		
≥10	7	2.7	0.82 (0.72-0.93)		
Total	150		1.00 (1.00-1.00)	0.99 (0.98-1.00)	1.00 (1.00-1.00)

P, prevalencia; (95% intervalo de confianza).

La característica morfológica del blastocisto que presentó más discrepancia entre laboratorios y expertos fue la valoración de la masa celular interna (MCI). La fiabilidad fue moderada según k (0.58 (0.36-0.79)) y buena según Gwet (0.72 (0.54-0.90)). Estas discrepancias se observaron especialmente, cuando los expertos consideraban que la MCI era de categoría D (0.70 (0.34-1.00)). Sin embargo, en la evaluación de la calidad del trofoectodermo se observan valores de fiabilidad más altos, que los obtenidos para la evaluación de la MCI, y un grado de acuerdo casi perfecto (0.94 (0.68-0.85)) entre laboratorios y el panel de expertos (Tabla 20).

Tabla 20. Acuerdo y fiabilidad entre laboratorios y el panel de expertos en la evaluación de la morfología de la masa celular interna (MCI) y el trofoectodermo.

Día 5	n	P	Acuerdo ponderado	Kappa	Gwet
MCI					
A	17	39.5	0.98 (0.96-1.00)		
B	12	27.9	0.89 (0.82-0.96)		
C	8	18.6	0.90 (0.79-1.00)		
D	6	14.0	0.70 (0.34-1.00)		
Total	43		0.90 (0.85-0.96)	0.58 (0.36-0.79)	0.72 (0.54-0.90)
Trofoectodermo					
A	9	25.0	0.94 (0.90-0.98)		
B	18	50.0	0.95 (0.90-1.00)		
C	3	8.3	0.93 (0.80-1.00)		
D	6	16.7	0.93 (0.87-0.98)		
Total	36		0.94 (0.68-0.85)	0.68 (0.51-0.86)	0.85 (0.76-0.94)

MCI, masa celular interna; P, prevalencia, (A, B, C, D), categorías de la evaluación de ASEBIR donde A tiene el mejor pronóstico y D el peor; (95% intervalo de confianza).

1.3. Destino del embrión

La fiabilidad entre laboratorios y el panel de expertos en el destino del embrión fue moderada y significativamente inferior a la observada en catalogación embrionaria (k: 0.51 vs 0.71 $p \leq 0.05$; Gwet: 0.51 vs 0.76 $p \leq 0.05$). Se observó una tendencia a un porcentaje de acuerdo entre laboratorios y expertos más bajo cuando el destino de un embrión para los expertos era crioconservar (transferir: 0.75; crioconservar: 0.56; descartar: 0.72) (Tabla 21)

.

Tabla 21. Acuerdo y fiabilidad entre laboratorios y el panel de expertos en la evaluación del destino final del embrión.

Destino final del embrión	n	P	Acuerdo crudo	Kappa	Gwet
Día 2					
Transferir	31	41.3	0.77 (0.62-0.93)		
Crioconservar	22	29.3	0.32 (0.11-0.53)		
Descartar	22	29.3	0.82 (0.65-0.99)		
Total	75		0.65 (0.54-0.76)	0.47 (0.32-0.63)	0.49 (0.32-0.65)
Día 3					
Transferir	30	40.0	0.73 (0.57-0.90)		
Crioconservar	21	28.0	0.57 (0.35-0.80)		
Descartar	24	32.0	0.67 (0.47-0.87)		
Total	75		0.67 (0.56-0.78)	0.49 (0.32-0.66)	0.51 (0.34-0.67)
Día 5					
Transferir	10	20.0	0.70 (0.40-1.00)		
Crioconservar	25	50.0	0.76 (0.58-0.94)		
Descartar	15	30.0	0.60 (0.33-0.87)		
Total	50		0.70 (0.57-0.83)	0.51 (0.29-0.73)	0.57 (0.38-0.76)
Global					
Transferir	71	35.5	0.75 (0.64-0.85)		
Crioconservar	68	34.0	0.56 (0.44-0.68)		
Descartar	61	30.5	0.72 (0.60-0.83)		
Total	200		0.67 (0.60-0.74)	0.51 (0.41-0.60)	0.51 (0.41-0.60)

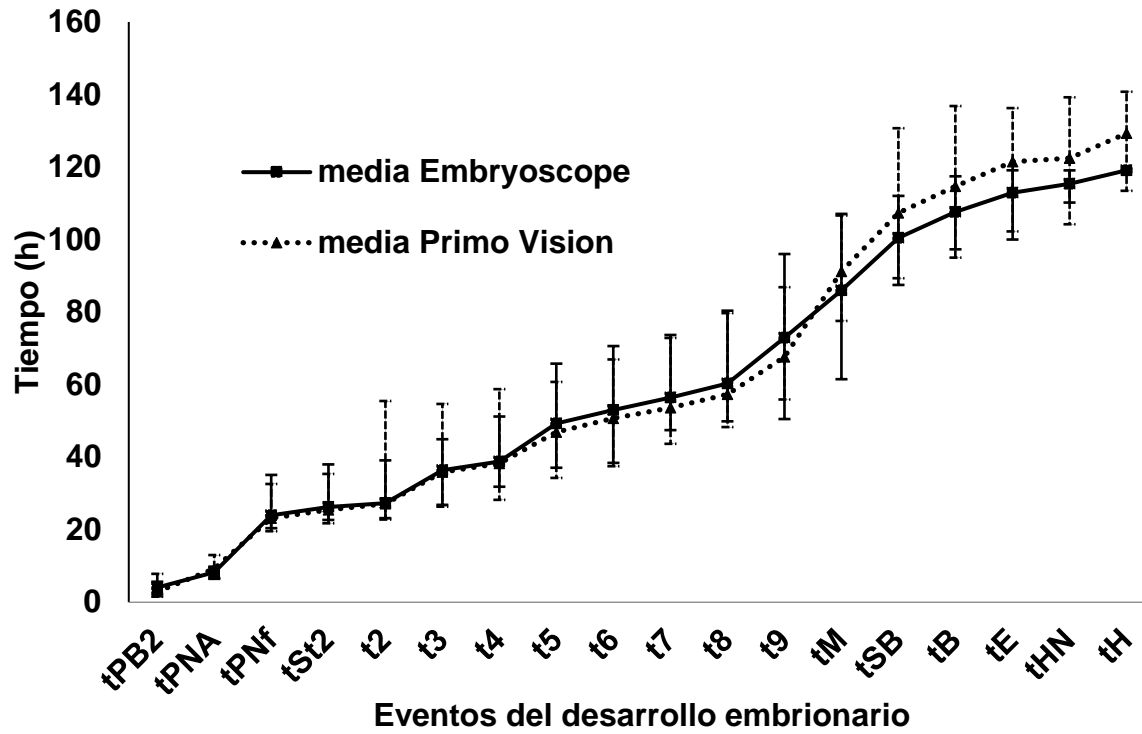
P, prevalencia; (95% intervalo de confianza).

2. Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante equipos *time-lapse*

2.1. Tiempos del desarrollo embrionaria en los equipos *time-lapse*

Cada uno de los eventos del desarrollo embrionario analizados con los equipos EmbryoScope™ y Primo Vision™ presentaron tiempos similares, aunque a partir del proceso de blastulación estos tiempos son más cortos en los embriones cultivados en EmbryoScope™ (Figura 20).

Figura 20. Tiempos medios de las variables cuantitativas para ambas plataformas de morfocinética, (máximos y mínimos).



Ambos equipos presentan dos periodos de mayor variabilidad entre laboratorios en los tiempos de los eventos del desarrollo. Un primer pico de variabilidad en los eventos iniciales del desarrollo embrionario (extrusión del segundo corpúsculo polar y aparición de pronúcleos) y un segundo pico de variabilidad entre el estadio de 8 células y el estadio de mórula (Figuras 21 y 22).

Figura 21 Media de los coeficientes de variación y sus correspondientes intervalos de confianza al 95% obtenidos en los diferentes tiempos del desarrollo embrionario desde la ICSI (inyección intracitoplasmática de espermatozoide) hasta el estadio de blastocisto, para ambas plataformas de morfocinética.

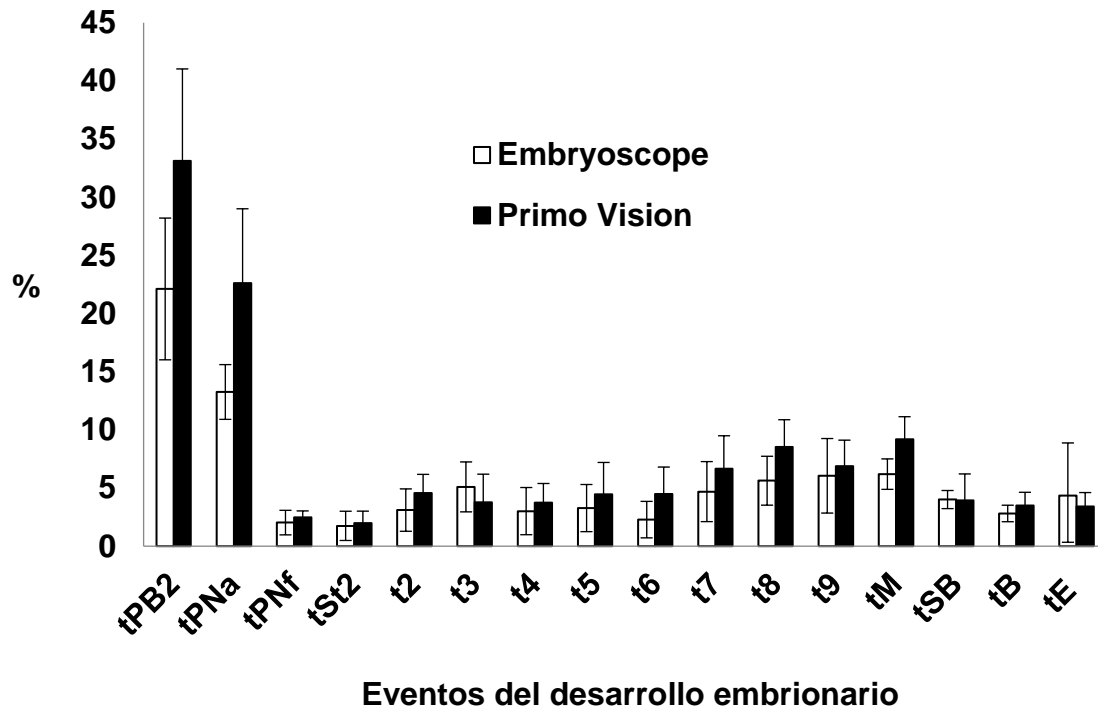
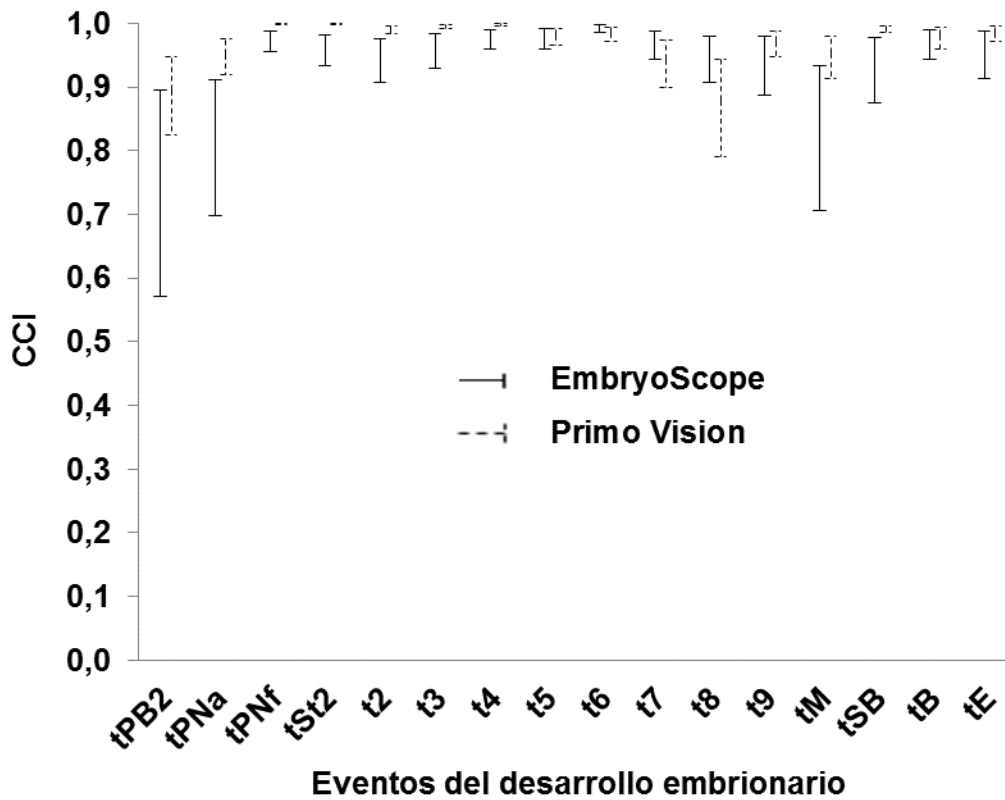


Figura 22. Coeficiente de correlación intraclass y su correspondiente intervalo de confianza al 95% obtenido a diferentes tiempos del desarrollo embrionario desde la ICSI (inyección intracitoplasmática de espermatozoide) hasta el estadio de blastocisto, para ambas plataformas de morfocinética.



2.2. Catalogación embrionaria

La variabilidad entre laboratorios en clasificación embrionaria según los criterios de ASEBIR fue similar (moderada-sustancial) tanto en los laboratorios que utilizaban la evaluación morfológica embrionaria clásica como aquellos que participaban con EmbryoScope™ en día 2, 3 y 5 del desarrollo embrionario. Sin embargo, esta variabilidad fue más alta (baja-moderada) entre los usuarios de Primo Vision™ (Tabla 22).

Tabla 22. Variabilidad entre laboratorios (Gwet) en catalogación embrionaria según criterios ASEBIR y destino clínico.

	Primo Vision™	Embryoscope™	EMEC
Catalogación			
Día 2	0.38 (0.24-0.52)	0.60 (0.47-0.72)	0.64 (0.52-0.75)
Día 3	0.39 (0.24-0.54)	0.56 (0.43-0.69)	0.63 (0.49-0.78)
Día 5	0.59 (0.39-0.79)	0.73 (0.60-0.86)	0.70 (0.62-0.78)
Destino			
Día 5	0.81 (0.71-0.90)	0.87 (0.80-0.94)	0.63 (0.54-0.72)

EMEC, Evaluación morfológica embrionaria clásica; (intervalo de confianza al 95%).

Similares conclusiones obtuvimos al analizar el porcentaje de embriones que presentaban un alto grado de acuerdo. En los laboratorios que usaron EMEC (30.4%, 24/79) y EmbryoScope™ (35.2%, 25/71) se obtuvo un porcentaje de embriones con alto grado de acuerdo similar. Este porcentaje fue en ambos casos superior al porcentaje de embriones con alto grado de acuerdo que se obtuvo entre los usuarios de Primo Vision™ (20%, 17/85).

Además, el porcentaje de embriones que presentaron un alto grado de acuerdo entre laboratorios cuando fueron clasificados según criterios ASEBIR, tanto con EMEC como con equipos *time-lapse* fue significativamente superior cuando los embriones fueron clasificados como de muy baja calidad (tipo D) por la mayoría de los laboratorios participantes, que para las categorías A, B o C (Tabla 23).

Tabla 23. Porcentaje de embriones con un alto grado de acuerdo en evaluación morfológica embrionaria según criterios ASEBIR.

	Primo Vision™	Embryoscope™	E MEC
A	18% (5/28)	33% (7/21)	37% (7/19)
B	0% (0/14)	8% (1/13)	9% (2/23) ^d
C	0% (0/18)	19% (4/21)	24% (6/25) ^f
D	48% (12/25) ^{a,b,c}	81% (13/16) ^{a,b,c,e}	75% (9/12) ^{a,b,c}

E MEC, Evaluación morfológica embrionaria clásica; (número de embriones con alto grado de acuerdo/número de embriones de esa categoría).

a $p \leq 0.05$ A vs D;

b $p \leq 0.05$ B vs D;

c $p \leq 0.05$ C vs D;

d $p \leq 0.05$ A vs B;

e $p \leq 0.05$ Primo Vision vs Embryoscope;

f $p \leq 0.05$ Primo Vision vs E MEC

2.3. Características morfológicas embrionarias

En la mayoría de las variables cualitativas del desarrollo embrionario, se observó una muy baja variabilidad, tanto al utilizar la evaluación morfológica embrionaria clásica como al usar los equipos EmbryoScope™ o Primo Vision™. Sin embargo, se observó una variabilidad más alta en: falsa división (Gwet: 0.69 (0.53-0.84) para usuarios de Primo Vision™ y 0.77 (0.62-0.93) para usuarios de EmbryoScope™) y presencia de vacuolas (Gwet: 0.58 (0.41-0.76) para usuarios de Primo Vision™ y 0.57 (0.36-0.79) para usuarios de EmbryoScope™) en usuarios de equipos *time-lapse* y para multinucleación, concretamente en los laboratorios que participaron con Primo Vision™ (Gwet: 0,70 (0.61-0.79)). En la evaluación de la asimetría de las blastómeras la variabilidad observada entre laboratorios fue alta en todos los métodos analizados (Tabla 24).

Tabla 24. Variabilidad entre laboratorios (Gwet) en la evaluación de variables embrionarias en los tres métodos de evaluación analizados.

	Primo Vision™	Embryoscope™	EMEC
Número de PN	0.99 (0.99 a 1)	0.99 (0.98 a 1)	
Tamaño de los PN	0.92 (0.86 a 0.97)	0.97 (0.90 a 1)	
Oposición de los PN	0.89 (0.83 a 0.94)	0.92 (0.86 a 0.97)	
División directa 1 a 3 células	0.87 (0.78 a 0.97)	0.88 (0.77 a 0.99)	
División directa 3 a 5 células	0.83 (0.74 a 0.93)	0.90 (0.81 a 1)	
Fragmentación	0.83 (0.78 a 0.88)	0.91 (0.87 a 0.95)	0.95 (0.92 a 0.98)
Multinucleación en estadio de 2 a 4 células	0.70 (0.61 a 0.79)	0.94 (0.90 a 0.97)	0.92 (0.86 a 0.98)
Falsa división	0.69 (0.53 a 0.84)	0.77 (0.62 a 0.93)	
Vacuolas	0.58 (0.41 a 0.76)	0.57 (0.36 a 0.79)	0.99 (0.99 a 1)
Asimetría	0.39 (0.22 a 0.55)	0.40 (0.25 a 0.54)	0.54 (0.44 a 0.64)

PN, pronúcleos; (Intervalo de confianza al 95%).

2.4. Destino del embrión

Respecto al destino del embrión, en día 5, el nivel de variabilidad fue inferior entre los usuarios de equipos *time-lapse* (Gwet: 0.81 (0.71-0.9) para usuarios de Primo Vision™ y 0.87 (0.80 a 0.94) para usuarios de EmbryoScope™) que entre los laboratorios que utilizaron EMEC (Gwet: 0.63 (0.54-0.72)) (Tabla 22). Además, el porcentaje de embriones con alto grado de acuerdo según su destino fue significativamente superior entre los usuarios de equipos *time-lapse* (>62%) que entre los usuarios de EMEC (25%) (Tabla 25).

EL porcentaje de embriones que presentaron un alto grado de acuerdo entre laboratorios respecto a su destino, tanto entre usuarios de EMEC como de equipos *time-lapse*, fue superior para embriones a descartar que en aquellos embriones a transferir (Tabla 25).

Tabla 25. Porcentaje de embriones con alto grado de acuerdo en decisión clínica.

	Primo Vision™	EmbryoScope™	EMEC
Transferir	0% (0/3)	67% (2/3)	33% (1/3)
Crioconservar	42% (5/12)	60% (9/15)	20% (3/15)
Descartar	85% (17/20)	100% (13/13)	50% (1/2)
Total	62% (22/35) ^a	77% (24/31) ^b	25% (5/20)

EMEC, Evaluación morfológica embrionaria clásica; (número de embriones con alto grado de acuerdo/número de embriones de esa categoría);

a $p \leq 0.05$ Primo Vision vs EMEC

b $p \leq 0.05$ EmbryoScope vs EMEC

3. Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad

De manera global e independientemente del producto analizado En la evaluación de fungibles en el programa de CCExt de CT se observó que la S fue superior a la E, y que el VPN fue superior que el VPP.

Al comparar los resultados según el material utilizado, observamos de manera global que se obtienen parámetros de rendimiento diagnóstico similares con ambos materiales, excepto en el VPP que resultó significativamente más alto al utilizar puntas de pipeta que medios de cultivo (0,82 [0,77-0,87] vs 0,71 [0,66-0,76], $p \leq 0.05$) (Tabla 26).

Tabla 26. Sensibilidad, especificidad, valor predictivo positivo y negativo y acuerdo global obtenidos en la evaluación de puntas y medios.

	Puntas de pipeta	Medios de cultivo
Sensibilidad	0.93 (0.89-0.97)	0.87 (0.83-0.91)
Especificidad	0.79 (0.73-0.85)	0.76 (0.71-0.81)
Valor predictivo negativo	0.92 (0.88-0.96)	0.90 (0.87-0.93)
Valor predictivo positivo	0.82 (0.77-0.87)	0.71 (0.66-0.76)
Acuerdo global	0.86 (0.83-0.90)	0.81 (0.77-0.84)

(Intervalo de confianza al 95%)

Se comparan los resultados obtenidos en este trabajo con los resultados de otros programas de CCExt de CT.

Comparando los resultados obtenidos analizando medios de cultivo en el CCExt de CT de ASEBIR y AAB (AAB, 2017) se observa, de manera global, parámetros de rendimiento diagnóstico inferiores en el primer programa a los obtenidos por los centros participantes en el CCExt de CT organizado por AAB (Tabla 27).

Al comparar los resultados obtenidos analizando materiales en el CCExt de CT de ASEBIR y los datos publicados por Castilla et al., (2010), se observa de manera global una tendencia a obtener parámetros de rendimiento diagnóstico superiores a los obtenidos en este último trabajo (Tabla 28).

Tabla 27. Comparación de sensibilidad, especificidad, valor predictivo positivo y negativo y acuerdo global entre los programas de CCExt de CT con medios de cultivo de ASEBIR y AAB (AAB, 2017).

	ASEBIR	AAB
Sensibilidad	0.87* (0.83-0.91)	0.97 (0.96-0.98)
Especificidad	0.76* (0.71-0.81)	0.94 (0.93-0.98)
Valor predictivo negativo	0.90* (0.87-0.93)	0.97 (0.96-0.98)
Valor predictivo positivo	0.71* (0.66-0.76)	0.94 (0.93-0.98)
Acuerdo global	0.81* (0.77-0.84)	0.96 (0.95-0.96)

* $p \leq 0.05$

Tabla 28. Comparación de sensibilidad, especificidad, valor predictivo positivo y negativo y acuerdo global entre los programas de CCExt de CT con materiales de ASEBIR (2013-16) y ASEBIR (2003-07; Castilla et al., 2010).

	ASEBIR (2013-16)	ASEBIR (2003-07)
Sensibilidad	0.93 (0.89-0.97)	0.83 (0.75-0.91)
Especificidad	0.79 (0.73-0.85)	0.68 (0.59-0.77)
Valor predictivo negativo	0.92 (0.88-0.96)	0.84 (0.76-0.92)
Valor predictivo positivo	0.82 (0.77-0.87)	0.67 (0.59-0.77)
Acuerdo global	0.86* (0.83-0.90)	0.75 (0.68-0.81)

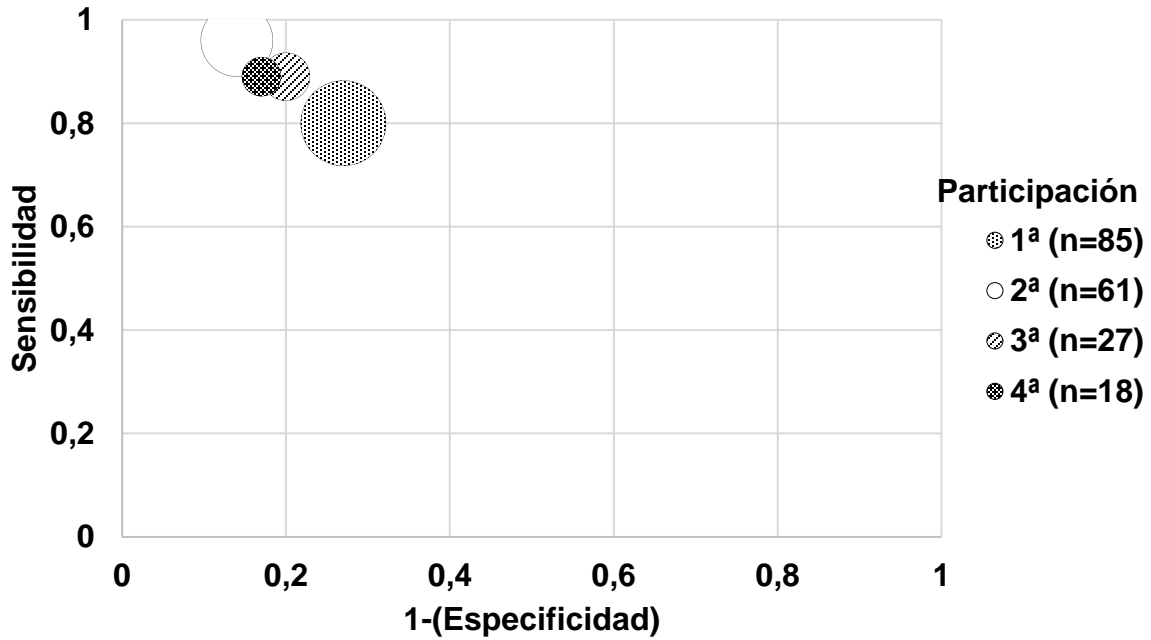
* $p \leq 0.05$

3.1. Influencia de la experiencia de los laboratorios en la participación en un programa de control de calidad externo de citotoxicidad

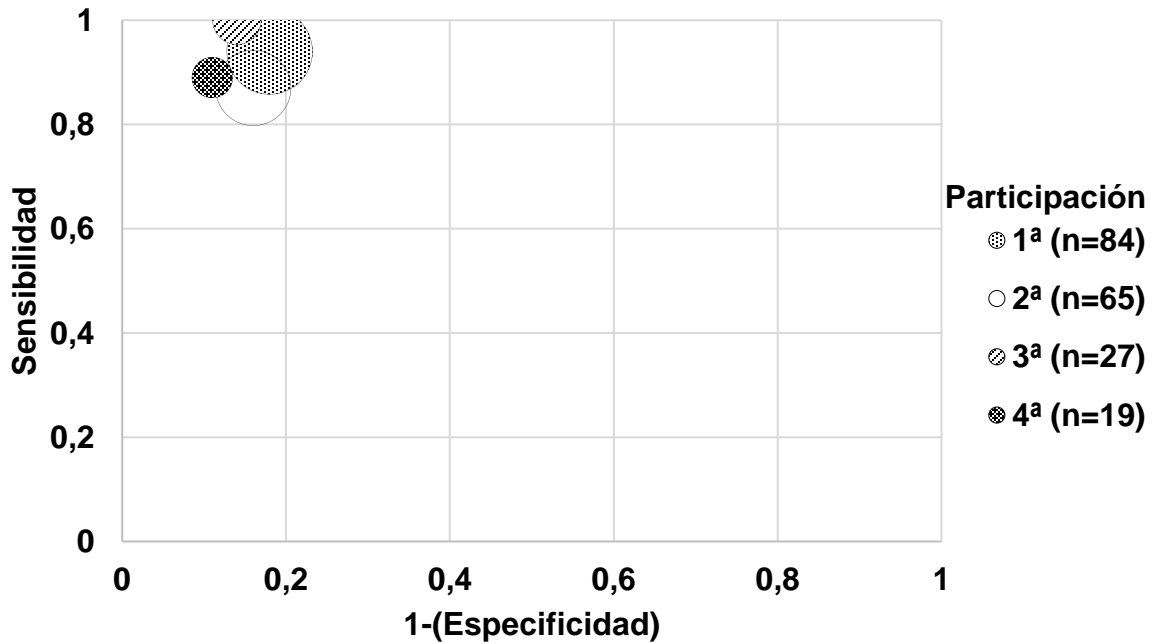
Al analizar las respuestas de los laboratorios por número de participaciones en el programa de CCExt de CT de ASEBIR, se observó que en la primera participación evaluando medios de cultivo, los resultados obtenidos son más bajos en S y E, que los obtenidos en la primera participación evaluando puntas de pipeta (Figura 23). Sin embargo, en las siguientes participaciones, con más experiencia, los resultados en sensibilidad y especificidad son similares entre ambos productos (Figura 23).

Figura 23. Espacio ROC obtenido en la valoración de medios (A) y puntas (B) según el número de participación en el CCExt de CT. El tamaño de los círculos se corresponde con el número de laboratorios.

A



B

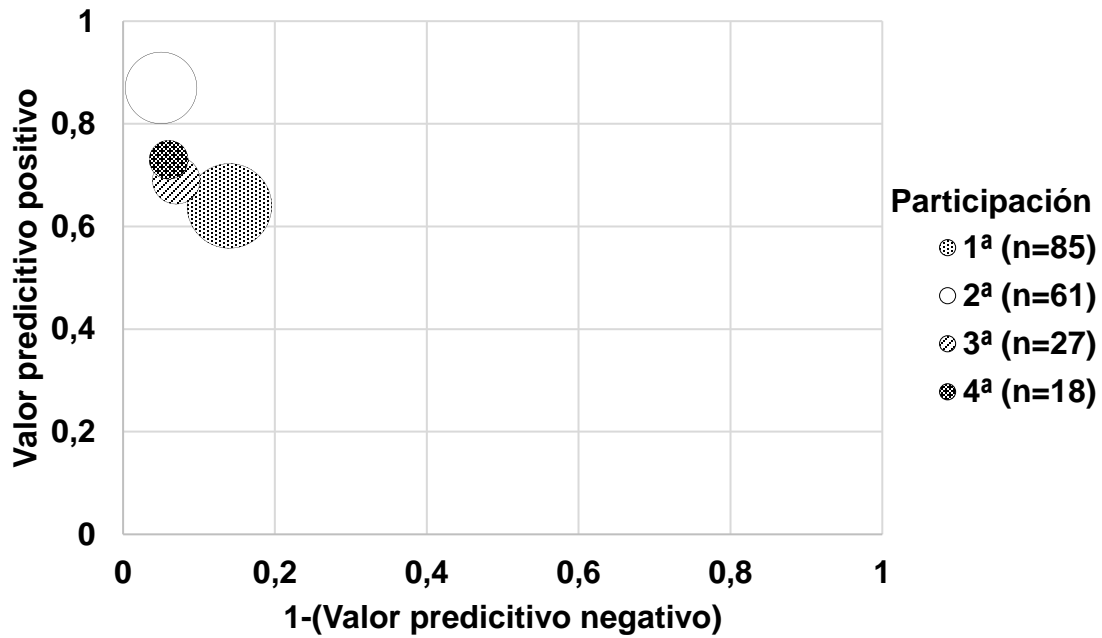


El VPP obtenido en la evaluación de puntas de pipeta fue superior al de medios tanto en la primera participación del laboratorio como en las participaciones en las que el laboratorio tenía más experiencia. Además, el VPP permanece constante y siempre superior a 0.8 en la evaluación de puntas de pipeta, al contrario de lo que ocurre en la evaluación de medios de cultivo.

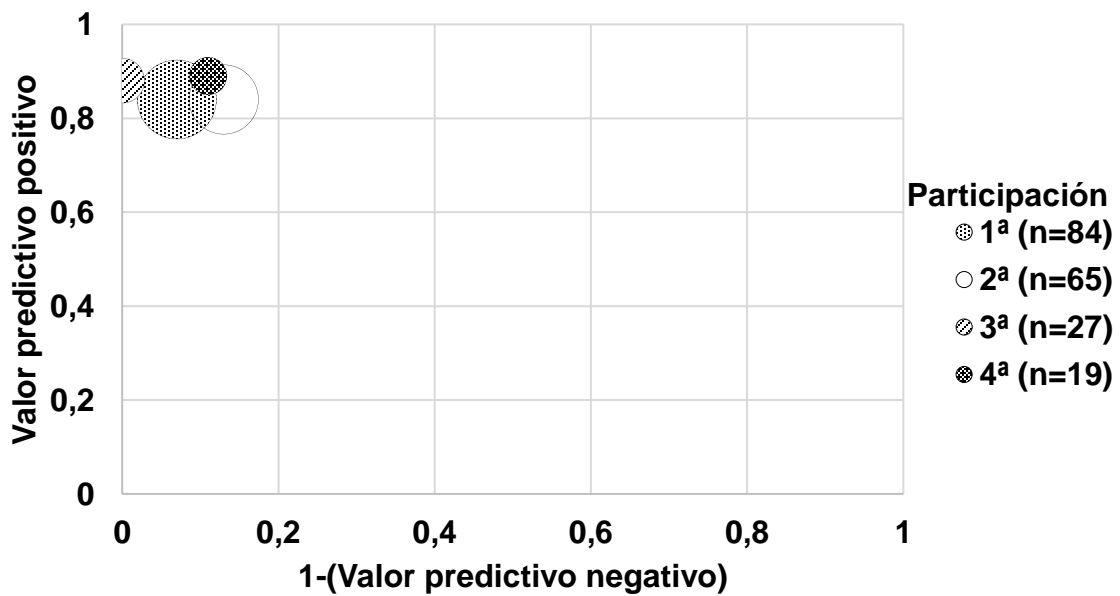
El VPN fue similar entre ambos productos, independientemente del número de participaciones en el CCExt de CT (Figura 24).

Figura 24. Representación gráfica del valor predictivo positivo frente al valor predictivo negativo obtenidos en la valoración de medios (A) y puntas (B) según el número de participación en el CCE-CT. El tamaño de los círculos se corresponde con el número de laboratorios.

A



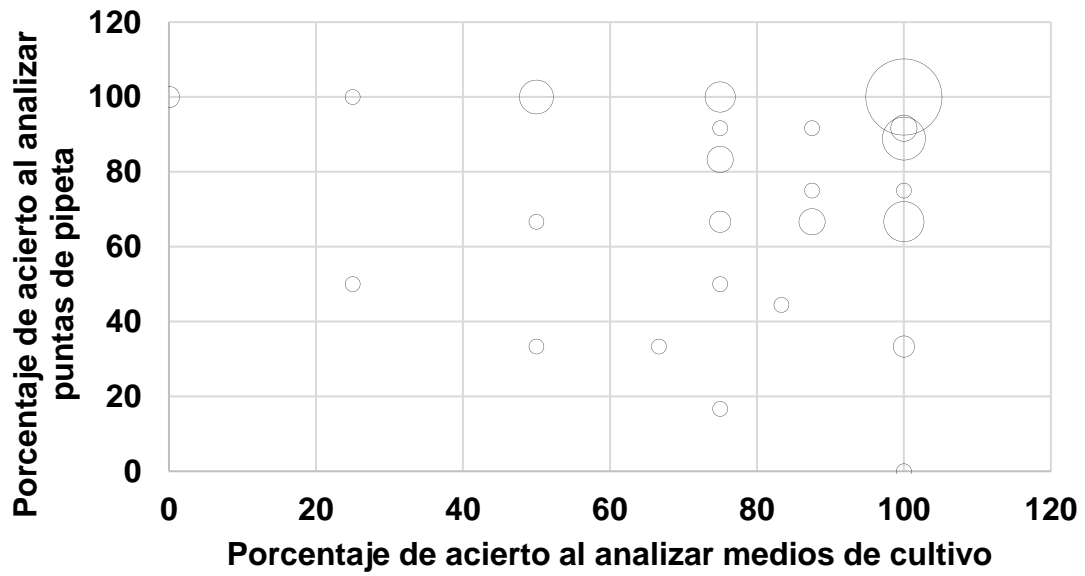
B



3.2. Correlación entre el grado de acuerdo en medios de cultivo y puntas de pipeta en el test de citotoxicidad

No se observó una correlación significativa entre el porcentaje de aciertos de un laboratorio en medios de cultivo y el de aciertos en puntas de pipeta ($r=0.026$) (Figura 25).

Figura 25. Porcentaje de acierto en puntas de pipeta frente a porcentaje de acierto en medios de cultivo en el CCE-CT. El tamaño de los círculos se corresponde con el número de laboratorios.



Discusión

Las técnicas de reproducción asistida son fundamentales para atender los deseos reproductivos de las parejas estériles y de las mujeres sin pareja masculina, para evitar la transmisión de enfermedades infecciosas y hereditarias o para preservar la fertilidad. Estas técnicas se organizan en tres áreas: análisis de gametos, cigotos y embriones, procesamiento de gametos y embriones y crioconservación de gametos y embriones (Castilla y Mantilla, 2016).

Como cualquier laboratorio clínico, el LRHA debe estar orientado a satisfacer las necesidades de sus usuarios mediante procesos de probada seguridad y eficiencia. Para ello es clave la implantación de sistemas de calidad y programas de CCInt y CCExt. Actualmente, diferentes sociedades científicas han recomendado la participación en programas de CCExt de EME y de CT, ya que permiten conocer la fiabilidad de los laboratorios y aumentar su eficiencia, respectivamente.

Respecto a los primeros, trabajos previos de nuestro grupo demostraron la importancia de utilizar valores de referencia basados en la opinión de expertos y pusieron de manifiesto las importantes diferencias existentes entre laboratorios y expertos en evaluación morfológica embrionaria clásica (EMEC). Con el objeto de profundizar en el estudio de estas discrepancias, desde 2012 hasta la actualidad se incorporó al programa de CCExt de EME de ASEBIR, el uso de la catalogación embrionaria publicada por ASEBIR poco tiempo antes.

En este tiempo, se han incorporado a la práctica diaria de los LRHA los equipos *time-lapse* para realizar la evaluación embrionaria, por lo que como ya ocurre para la EMEC, es clave el desarrollo de programas de CCExt que evalúen y validen su utilidad, especialmente por su reciente implantación (Harper et al., 2012).

Por otra parte, y para garantizar la viabilidad de gametos y embriones, el LRHA debe tener la capacidad de identificar los fungibles tóxicos antes de su uso, para lo cual se recomienda testar la CT de estos fungibles previo a su utilización. Los programas de CCExt de CT en los LRHA están implantados desde hace tiempo, pero es necesario su estandarización ya que los fungibles que se evalúan son muy diversos (puntas de pipeta, placas de cultivo, catéteres de transferencia, medios de cultivo, etc.) y se desconoce si el rendimiento de un laboratorio en la evaluación de fungibles tóxicos depende del tipo de fungible evaluado.

En este trabajo, nos proponemos abordar diferentes aspectos de los programas de CCExt del LRHA, utilizando una metodología estandarizada como son las recomendaciones para publicar estudios de fiabilidad y acuerdo (GRRAS).

1. Acuerdo y fiabilidad entre laboratorios y un panel de expertos en evaluación morfológica embrionaria clásica

El análisis de los resultados obtenidos en el programa de CCExt de ASEBIR desde 2012, utilizando una catalogación estandarizada, ha demostrado el alto grado de acuerdo existente entre laboratorios y expertos en la evaluación de embriones de buena calidad. Este grado de acuerdo es similar al observado en este mismo programa previo a la incorporación de la catalogación estandarizada (Ruiz de Assin et al., 2009), utilizando un sistema de catalogación embrionaria de tres categorías ("bueno", "intermedio" y "pobre"). Sin embargo, a diferencia de estos autores, observamos una tendencia a un mayor desacuerdo entre laboratorios y el panel de expertos en los embriones tipo D, posiblemente, debido a que en la asimilación de

un sistema de catalogación estandarizado de cuatro categorías se presentan más dificultades en la definición de embriones de baja calidad que en los de calidad óptima. En consecuencia, los usuarios de la catalogación de ASEBIR, analizados en este trabajo, podrían haber relacionado "buena" calidad con A, "intermedio" con B y "pobres" con C o D. Esto explicaría por qué, en nuestros resultados, cuando el panel de expertos catalogó un embrión como D, la mayoría de los laboratorios lo clasificó en una categoría más alta.

Otros autores, que han analizado fiabilidad y acuerdo en evaluación morfológica embrionaria (Arce et al., 2006; Baxter Bendus et al., 2006; Paternot et al., 2009; Storr et al., 2017) no analizan el acuerdo según las categorías embrionarias. Sin embargo, este análisis es importante, ya que la inexactitud en la catalogación concreta de embriones de mala calidad podría llevar a una menor especificidad en la selección de embriones, provocando que embriones con menor potencial de implantación se transfieran antes que otros con mayor potencial. Esto generaría un período de espera más largo para lograr el embarazo aumentando el stress en las parejas que requieren de técnicas de reproducción asistida (Gianaroli et al., 2013).

Además de tener en cuenta las categorías embrionarias, se debe tener en cuenta que los estadísticos de fiabilidad y acuerdo en estudios basados en esquemas de catalogación de tres o más categorías deben ser ponderados, como se realiza en este trabajo, ya que errores en una sola categoría (por ejemplo, clasificar un embrión D como C) tienen menor significación clínica que los errores

que implican diferencias en más de una categoría (por ejemplo, clasificar un embrión D como B o A).

La fiabilidad “casi perfecta” obtenida en la evaluación de la fragmentación embrionaria, es superior a la publicada hasta el momento (Arce et al., 2006; Paternot et al., 2009, 2011). La mayor fiabilidad obtenida en comparación con la obtenida por Arce et al. (2006), puede ser debido al tipo de imágenes analizadas. En este último estudio, la evaluación de un panel de expertos se basó en imágenes 2D, mientras que la evaluación realizada por el resto de laboratorios fue sobre imágenes en 3D. Por el contrario, en nuestro estudio tanto los laboratorios como el grupo de expertos tuvieron acceso a imágenes en 3D enfocadas en forma de barrido a lo largo de todo el embrión. Esta diferencia podría afectar a los niveles de fiabilidad y de acuerdo, ya que las imágenes 2D no permiten al observador determinar, por ejemplo, el volumen total del embrión que es ocupado por fragmentos. Otra causa, podría ser el diferente número de categorías utilizadas en ambos estudios para valorar la fragmentación. En nuestro estudio se utilizan dos categorías para valorar fragmentación ($\leq 10\%$ y $> 10\%$) mientras que en el estudio de Arce et al. (2006) se utilizan seis categorías para evaluar la misma característica, lo que podría llevar a un menor grado de fiabilidad al analizar fragmentación (Matava et al., 2015).

En la evaluación de la asimetría de las blastómeras, la fiabilidad obtenida en nuestro estudio fue significativamente superior a la obtenida en los estudios multicéntricos publicados (Arce et al., 2006; Paternot et al., 2011).

Al analizar presencia/ausencia de multinucleación o de vacuolas en embriones en estadio de células y la catalogación de blastocistos entre laboratorios y un panel de expertos mediante EMEC se obtuvo un porcentaje de acuerdo ponderado casi total ($>0,92$) entre laboratorios y el panel de expertos, así como una elevada fiabilidad (Gwet: $>0,90$), mientras que al utilizar el coeficiente kappa (k) para medir la fiabilidad, este fue moderado, inferior a 0,7. Esta paradoja de obtener elevados porcentajes de acuerdo con k bajas, ya comentada en la introducción (6.3.1. Parámetros de fiabilidad para variables categóricas (nominales, dicotómicas y ordinales)) es debida a grandes diferencias en la prevalencia de las categorías de los embriones analizados. La probabilidad de acuerdo debida solamente al azar se ve aumentada en poblaciones muy homogéneas, es decir, en poblaciones en las que la presencia de una característica tiene alta o baja prevalencia, como es el caso, por ejemplo, de la presencia de vacuolas o multinucleación (que tienen baja prevalencia) (Cicchetti y Feinstein, 1990; Feinstein y Cicchetti, 1990; Sim y Wright, 2005; Quemadura y Weir, 2011). Al evaluar estas características, los embriólogos pueden estar de acuerdo, por casualidad, en que la mayoría de los embriones no presentan multinucleación; sin embargo, este acuerdo no ofrecería ninguna garantía de que los embriólogos pudieran identificar, de forma habitual, estas alteraciones en el embrión. De hecho, nuestros resultados, muestran un alto grado de acuerdo para los embriones que no presentan las anomalías (multinucleación y vacuolas) y bajo grado de acuerdo para los embriones que si las presentaban.

Por consiguiente, los programas de CCEst de evaluación embrionaria deben incluir entre los embriones a analizar, una mayor proporción de embriones que

presentan anomalías morfológicas, lo que permitiría a los laboratorios mejorar su capacidad de identificar estas anomalías.

Nuestros resultados muestran que la fiabilidad entre laboratorios y el panel de expertos sobre el número de células tiende a disminuir de forma inversamente proporcional al número de blastómeras. Esto se debería al lógico aumento de la dificultad de identificar células conforme aumenta el número de estas. Esta idea, coincide con investigaciones previas (Arce et al., 2006; Paternot et al., 2011). Sin embargo, nuestros valores de fiabilidad son superiores a los publicados por Arce et al. (2006), debido probablemente, a la forma de analizar las imágenes en 2D en este último estudio, como se comentó anteriormente.

A pesar de las limitaciones del coeficiente k , los resultados obtenidos de fiabilidad y acuerdo en catalogación de embriones son superiores a los publicados anteriormente por otros autores usando este coeficiente (Arce et al., 2006; Baxter Bendus et al., 2006; Paternot et al., 2009; Storr et al., 2017). Esto a pesar de que en nuestro trabajo se utilizó una clasificación de más categorías (cuatro categorías) que tiende a producir valores más bajos de k (Matava et al., 2015). La discrepancia, entre nuestros datos y los publicados por otros autores podría deberse al diferente diseño de nuestro estudio (en el que la categoría elegida por la mayoría de laboratorios se comparó con la asignada por un panel de expertos) en comparación con el diseño empleado por otros investigadores (variabilidad interobservador). Sin embargo, el hecho de que la detección de multinucleación o vacuolas sea deficiente cuando están presentes, en comparación al alto grado de acuerdo observado en la determinación de la ausencia de multinucleación o vacuolas, podría explicar

parcialmente la elevada variabilidad entre observadores, publicada por otros autores, cuando se ha analizado la fiabilidad en la evaluación de estas características sin diferenciar si estaban o no presentes (Arce et al., 2006; Paternot et al., 2009, 2011).

En la catalogación de blastocisto, se observan altos valores globales de fiabilidad y acuerdo. Sin embargo, observamos al igual que Storr et al., (2017), niveles bajos de fiabilidad entre observadores en la evaluación de la MCI. Estos autores no analizaron la influencia de la calidad de la MCI en esta baja fiabilidad. Nosotros, al tener en cuenta este factor, observamos que la baja fiabilidad obtenida al analizar la MCI es debida principalmente a las MCI de mala calidad. El menor grado de acuerdo obtenido al evaluar las MCI tipo D, no parece afectar a la clasificación general de blastocistos, esto podría ser debido al mayor peso asignado a la calidad del trofoectodermo en la clasificación ASEBIR de blastocistos (Hurtado de Mendoza, 2015).

Con respecto al destino final del embrión, la baja fiabilidad y acuerdo observados, especialmente en la decisión de crioconservar embriones, en comparación a los valores de fiabilidad y acuerdo encontrados en catalogación embrionaria, podría deberse a las diferentes políticas de transferencia y criopreservación propias de cada centro. Mientras que en unos la política de transferencia de embrión único puede estar completamente implementada, en otros será más habitual no realizar transferencia en fresco y congelar todo (freeze all). Estas consideraciones podrían explicar la alta variabilidad en las tasas de embarazo derivado de criotransferencia observado entre laboratorios por el ESHRE IVF

Monitoring Consortium, en comparación con la variabilidad observada tras transferencia de embriones frescos (coeficiente de variación: 0,41 vs 0.18, respectivamente) (European IVF-Monitoring Consortium for the European Society of Human Reproduction and Embryology et al., 2016). Creemos que deben realizarse estudios más profundos para comparar las políticas de transferencia existentes con el fin de lograr niveles más altos de acuerdo respecto al destino final del embrión especialmente cuando los embriones deberían ser criconservados.

La relativa poca fiabilidad observada en el destino del embrión en día 5, en comparación con lo publicado por Storr et al. (2017), podría ser debido a que nuestro programa considera tres opciones de respuesta (transferencia, criopreservar o descartar) mientras que Storr et al. (2017) consideran sólo dos (transferencia o no de transferencia).

Una de las limitaciones de nuestro trabajo es que no disponemos de información sobre el nivel de experiencia de los embriólogos que participan en el CCEExt. Sin embargo, Baxter Bendus et al (2006) no observaron relación entre el nivel de formación o de los años de experiencia de los embriólogos participantes y el acuerdo entre observadores al evaluar diferentes parámetros de una clasificación morfológica embrionaria. Otra limitación de nuestro estudio es que el acuerdo intraobservador no fue analizado (el grado en el que un observador consigue idénticos resultados de forma repetida bajo condiciones de evaluación similares). Sin embargo, varios autores han publicado altos niveles de acuerdo intraobservador en situaciones similares (Arce et al., 2006; Baxter Bendus et al., 2006). Otra posible limitación es que la participación en el programa de CCEExt de ASEBIR es

voluntaria, esta circunstancia podría generar un sesgo de selección, en el que sólo los embriólogos con un interés particular en mantener altos estándares de calidad en sus laboratorios participarían en el CCEExt, mientras que otros, con menos interés, podrían optar por no hacerlo. Una limitación que afecta a todos los programas de CCEExt de evaluación morfológica embrionaria es que dichos programas no tienen en cuenta otras características relevantes para la toma de decisiones en el laboratorio de embriología, como la historia clínica de los pacientes o las condiciones de cultivo.

En conclusión, los valores más altos de fiabilidad y acuerdo en evaluación morfológica embrionaria entre un panel de expertos y los laboratorios se encuentran en la evaluación de los embriones sin anomalías morfológicas. Cuando se observan diferencias entre el panel de expertos y los laboratorios, se deben a la presencia de vacuolas y multinucleación, y MCI de baja calidad. Consideramos necesario desarrollar actividades para mejorar la evaluación de estas características morfológicas. Estos resultados deben tenerse en cuenta para diseñar mejores esquemas de CCEExt.

2. Variabilidad entre laboratorios en la evaluación morfológica embrionaria mediante *time-lapse*

Los equipos *time-lapse* permiten la obtención de más información sobre el desarrollo embrionario y dan la posibilidad de visualizar las imágenes tantas veces como sea necesario, sin perjudicar las condiciones de cultivo embrionario, lo que debería facilitar la obtención de resultados precisos. Sin embargo, según Lundin y

Ahlstrom (2015), el hecho de que los eventos del desarrollo embrionario puedan ser definidos en diferentes momentos (inicial, medio o final), podría dificultar la comparación de los resultados obtenidos por diferentes laboratorios. Nuestros resultados, tras analizar los datos de un ensayo piloto de CCExt de EME mediante *time-lapse*, parecen confirmar esta hipótesis, observándose un pico de variabilidad entre laboratorios al analizar eventos que tardan más tiempo en producirse (30-45 min.), como la extrusión del segundo corpúsculo polar o la aparición de los pronúcleos. Por el contrario, en eventos que se producen casi instantáneamente, como la división celular o la desaparición de los pronúcleos, la variabilidad observada entre laboratorios fue significativamente menor.

Otro pico de variabilidad entre laboratorios, observado con este ensayo piloto, tuvo lugar a partir del estadio de 8 células, posiblemente porque las células están más próximas entre sí haciendo más difícil su conteo y porque en estos momentos suele comenzar el proceso de compactación. Además, esta compactación puede ser parcial o incompleta, cuando alguna célula es excluida de la compactación o cuando el proceso de compactación no ha terminado, lo que hace imposible distinguir claramente, si estamos ante un embrión en el que pueden contarse células o ante uno que ha comenzado la etapa de mórula. Esta observación estaría relacionada con el descenso de acuerdo y fiabilidad en la evaluación del número de blastómeras mediante EMEC comentado anteriormente.

Tras la revisión sistemática realizada por Kirkegaard et al. (2013) sobre los parámetros morfocinéticos relacionados con la consecución de embarazo o de embriones euploides, estos autores llegaron a la conclusión de que los pocos

parámetros con capacidad predictiva, publicados por más de un laboratorio, estaban fuera de los períodos de alta variabilidad entre laboratorios comentados anteriormente. Esto sugiere que los parámetros para estos períodos de alta variabilidad entre laboratorios tienen poca validez externa y no deben ser utilizados en algoritmos de clasificación embrionaria. De hecho, uno de los algoritmos más utilizados, publicado por Meseguer et al. (2011) y actualizado por Basile et al. (2015), no incluye ninguno de los parámetros incluidos en estos picos de alta variabilidad entre observadores.

La baja variabilidad observada entre laboratorios, al analizar mediante *time-lapse* los eventos de morfocinética, coincide con la publicada por Sundvall et al. (2013). En nuestro estudio, la validez de esta conclusión se vio reforzada por la participación de 24 laboratorios, comparado con los tres observadores del mismo laboratorio estudiados por Sundvall et al. (2013)

Tras comprobar los resultados del ensayo piloto de EME mediante *time-lapse* con los obtenidos en el programa de CCExt de EMEC, no se observó una disminución en la variabilidad entre laboratorios usuarios de equipos *time-lapse* frente a los usuarios de EMEC, en la evaluación de ninguna característica morfológica ni en clasificación de embriones según los criterios ASEBIR. Este resultado es contrario a lo que habría cabido esperar de una menor variabilidad en usuarios de equipos *time-lapse*, y podría deberse en parte, a que muchas características morfológicas embrionaria son transitorias y los criterios para evaluar dichas características no están bien definidos actualmente para los equipos *time-lapse* (Hardarson et al., 2006; Chavez et al., 2012).

Entre los usuarios de Primo Vision™ la variabilidad fue significativamente mayor en la evaluación de la fragmentación y multinucleación que entre los usuarios de EmbryoScope™ y EMEC. Esto puede ser debido a limitaciones técnicas (ópticas) de esta plataforma, dificultando la evaluación de estas características. Este hallazgo corrobora las recomendaciones de otros autores (Harper et al., 2012; Lundin y Ahlstrom, 2015) de que cualquier nueva tecnología de selección de embriones debe ser evaluada antes de su incorporación a los laboratorios de reproducción asistida.

Cabe destacar que la ya comentada ausencia de mejora en la variabilidad entre observadores en los usuarios de *time-lapse* incluye la evaluación de la asimetría, una característica en la que podría esperarse un menor acuerdo entre los usuarios de EMEC, ya que estos no disponen de herramientas precisas para evaluar la asimetría de la célula. En cambio, muchos equipos *time-lapse* sí que disponen de estos instrumentos. Estos resultados podrían explicarse por tres factores. En primer lugar, no todos los laboratorios utilizan estos instrumentos. En segundo lugar, no todos los laboratorios emplean la misma definición de asimetría embrionaria y, por último, la mencionada naturaleza transitoria de las características morfológicas, incluyendo asimetría, puede dificultar su análisis (Aparicio-Ruiz et al., 2016).

Al contrario de lo que ocurre en catalogación embrionaria, la variabilidad en decisión clínica es menor entre usuarios que utilizan equipos *time-lapse* que entre aquellos que utilizan EMEC. Este puede deberse a la baja variabilidad observada en características morfológicas, que sólo pueden ser observados por *time-lapse*

(Kirkegaard et al., 2012, Ciray et al., 2014), y que han sido relacionadas con una baja calidad embrionaria (Rubio et al., 2012; Athayde Wirka et al., 2014; Liu et al., 2015). Sin embargo, la variabilidad obtenida en las características que pueden observarse con ambos métodos es más alta que en las características anteriores. Esto parece ser un resultado lógico, puesto que, en la toma de decisiones, los usuarios de EMEC, sólo utilizan las características para las cuales hay un menor acuerdo; por lo tanto, la decisión clínica en base a estas variables presenta una mayor variabilidad que cuando se tienen en cuenta más variables (como es el caso de los laboratorios que utilizan *time-lapse*).

Algunos autores han sugerido que el uso de algoritmos basados solo en eventos morfocinéticos debe complementarse con la observación de características morfológicas clásicas, ya sea con equipos *time-lapse* (Aparicio-Ruiz et al., 2016) o por EMEC (Kieslinger et al., 2016). Sin embargo, nuestros resultados muestran que esta práctica sigue presentando variabilidad entre laboratorios. El uso de la tecnología *time-lapse* para evaluar variables morfológicas clásicas requiere que estas variables estén claramente definidas. Por ejemplo, el grado de fragmentación puede evaluarse de varias formas: como el grado de fragmentación máxima alcanzada durante todo el desarrollo del embrión; como el grado de fragmentación en el último fotograma obtenido antes de cada ciclo celular; o como el grado de fragmentación que presenta el embrión el día de la transferencia (Ahlstrom et al., 2016). Según nuestros datos, el uso de equipos *time-lapse* no garantiza un mayor acuerdo entre laboratorios en la evaluación de variables morfológicas. Por lo tanto, creemos que antes de recomendar la incorporación de estas variables en los

algoritmos correspondientes, la definición de parámetros tales como la asimetría, multinucleation y vacuolas, utilizando tecnología *time-lapse*, debe ser estandarizada.

Los resultados de nuestro estudio muestran que los laboratorios de embriología tienden a presentar menor variabilidad en la evaluación de embriones de baja calidad que en los de alta calidad. De igual manera, tienden a presentar menor variabilidad en el destino clínico de los embriones que deben descartarse que en el de los que deben transferirse, independientemente de la técnica de evaluación embrionaria utilizada. Estos resultados coinciden con los publicados por Arce et al. (2006) y Castilla et al. (2010) para los usuarios de EMEC. El alto grado de acuerdo en la clasificación de embriones de baja calidad produciría una baja tasa de rechazo de embriones viables. Por otro lado, la heterogeneidad de los resultados observados en la clasificación de los embriones de alta calidad dificulta la obtención de modelos de clasificación reproducibles que permitan un incremento en las tasas de implantación (Kirkegaard et al., 2014).

Kirkegaard et al. (2015) sugirieron que las catalogaciones embrionarias deben utilizarse para clasificar en lugar de para seleccionar, para evitar el riesgo de desechar embriones viables, teniendo en cuenta que ningún modelo de catalogación ha demostrado un 100% de sensibilidad y especificidad en la predicción del potencial de implantación. De hecho, Mastenbroek et al. (2011), teniendo en cuenta este riesgo, cuestionó la necesidad de cualquier selección embrionaria y recomendaron, al menos, la redefinición de “selección de embriones”. Para minimizar este efecto negativo de la selección embrionaria, es crucial controlar

la exactitud de los criterios de selección de embriones, que, como muestran nuestros resultados, varía según la tecnología utilizada.

Existen distintas limitaciones en este estudio. Los embriones evaluados por cada método son diferentes y han estado sometidos a diferentes condiciones de cultivo. Estas diferencias en las condiciones de cultivo producirían diferencias en el desarrollo de los embriones y en su morfología, esto podría hacer que los laboratorios coincidan entre ellos a la hora de evaluar los embriones más con un método que con otro. Sin embargo, los tiempos de desarrollo de los embriones analizados fueron similares para cada una de las plataformas y aunque no hubo diferencias significativas, a partir de la etapa de mórula se observa un desarrollo embrionario más lento en los embriones examinados usando el sistema Primo Vision™. Esto puede ser debido a que los embriones en el sistema EmbryoScope™ fueron cultivados en una atmósfera al 5% de oxígeno, mientras que los cultivados en el sistema Primo Vision™ son cultivados en una atmósfera con un 21% oxígeno. Esto coincide con la idea de que embriones cultivados a bajas concentraciones de oxígeno presentan una velocidad de desarrollo más rápida que los embriones cultivados a concentraciones atmosféricas de oxígeno (De los santos et al., 2013; Kirkegaard et al., 2013).

Otra limitación es que se desconocen las tasas de implantación de los embriones evaluados y, por tanto, no se ha podido determinar, si la variabilidad entre laboratorios, observada en este estudio, en decisión clínica y en catalogación embrionaria puede reducir la efectividad de los ciclos de reproducción asistida. Además, en este estudio, no se analizó la variabilidad intraobservador, sin embargo,

Sundvall et al. (2013) publicaron que en una situación similar la variabilidad intraobservador es baja. Por último, no se puede descartar la posibilidad del sesgo de participación ya comentada en el estudio anterior.

En vista de los resultados de este estudio la variabilidad entre laboratorios en la determinación de parámetros morfocinéticos depende de la característica del evento, siendo alta si el evento analizado se produce lentamente. Creemos necesario desarrollar directrices para definir e interpretar las características morfológicas clásicas observadas por *time-lapse* antes de incorporar estas características a algoritmos diagnósticos basados en *time-lapse*.

3. Influencia del tipo de producto a analizar en un programa de control de calidad externo de citotoxicidad

La identificación de fungibles que puedan ser citotóxicos es esencial para el óptimo rendimiento de los LRHA (Lierman et al., 2007). Con este objetivo, es fundamental testar los fungibles a usar en el LRHA mediante bioensayos como el TSE (Nijs et al., 2009; OMS, 2010; ESHRE Guideline Group on Good Practice in IVF Labs et al., 2016). Para garantizar la validez de estos TSE se recomienda participar en programas de CCExt (De Jonge et al., 2003; Nijs et al., 2009; Castilla et al., 2010; Hossain et al., 2010; OMS, 2010). Los resultados obtenidos en estos programas deberían no verse influidos por el tipo de fungible utilizado. Sin embargo, nuestros resultados demuestran que, a la hora de evaluar el rendimiento de estos bioensayos, la capacidad para detectar productos tóxicos se relaciona con el tipo de fungible testado.

En nuestro estudio, se observa una tendencia a obtener resultados de S, E, VPP y VPN analizando puntas de pipeta, superiores a los publicados por Castilla et al. (2010). Esto podría deberse a que en los años incluidos en nuestro estudio, los laboratorios participantes solo utilizaron el TSE como test y solo analizan un tipo de material (puntas de pipeta), al contrario de lo que ocurría en el programa de CCExt cuando fue analizado por Castilla et al. (2010), en los que los laboratorios emplearon diferentes métodos de evaluación, lo que influye en los resultados de los bioensayos (Miller et al., 2001; Cai et al., 2006) y se analizaron diversos tipos de materiales tóxicos, lo que también podría influir en la detección de CT.

Los resultados en S, E, VPP y VPN obtenidos en nuestro estudio analizando medios de cultivo, son inferiores a los observados en el programa de CCExt de CT de la AAB que también utiliza medios de cultivo en su programa (AAB, 2017). Estas diferencias podrían deberse a diferencias en la manipulación de los medios de cultivo para convertirlos en tóxicos en los dos programas de CCExt de CT analizados.

Los laboratorios participantes detectan mejor los productos citotóxicos que los que no lo son, pues se observan VPN superiores a VPP. Desde un punto de vista clínico parece menos perjudicial equivocarse clasificando fungible no tóxico como tóxico que un fungible tóxico como no tóxico. Pues en este último caso estaríamos utilizando en nuestro laboratorio material perjudicial para los gametos. Esta tendencia coincide con la observada en nuestro programa años anteriores analizando materiales (Castilla et al., 2010) y con la observada en programas similares analizando medios de cultivo (AAB, 2017).

Los mismos laboratorios son más eficaces detectando toxicidad en puntas de pipeta que en medios de cultivo. Estas diferencias no son achacables a diferencias en el nivel de toxicidad de los productos utilizados pues los organizadores verificaron mediante un TSE que el nivel de toxicidad de los materiales producía siempre reducciones en la movilidad espermática a las 24h superiores al 50% de la inicial. Además, estas diferencias hubieran afectado principalmente al VPN, pues ante un producto muy tóxico se darían pocos falsos negativos (FN). Sin embargo, las diferencias entre productos se han observado en VPP, lo que solo puede deberse a que los TSE realizados por los laboratorios presentan falsos positivos (FP), es decir detectan toxicidad en fungibles que no han sido manipulados y no son tóxicos. Al observar en nuestros resultados, que las diferencias se mantienen, independientemente del número de veces que un laboratorio ha participado en el programa de control de calidad, no se pueden achacar estas, a una curva de aprendizaje.

Creemos más bien, que estas diferencias pueden estar relacionadas con el método utilizado por los laboratorios para evaluar la CT de los fungibles. Cuando se evalúan puntas de pipeta, los espermatozoides control se encuentran en el mismo medio de cultivo que los espermatozoides que están bajo la influencia de la toxicidad de las puntas de pipeta. Sin embargo, cuando se evalúan medios de cultivo los espermatozoides control se encuentran en un medio diferente al que se encuentran los espermatozoides incubados con el medio problema. Una manera de evitar esto, sería aconsejar a los laboratorios que para evaluar los medios incubaran durante un tiempo (unas horas) los espermatozoides con el medio problema y

posteriormente realizaran el cultivo *overnight* en el mismo medio de cultivo que los espermatozoides control. De esta manera, se evitaría que pequeñas diferencias en movilidad puedan achacarse a que el medio testado es tóxico, cuando en realidad se debe a la diferente composición de los medios de cultivo utilizados. Otra alternativa sería incluir en los envíos de medio de cultivo, una alícuota no manipulada del mismo medio de cultivo problema, para ser utilizada como control en TSE.

Otra posible causa de las diferencias observadas podría ser que las condiciones de transporte, almacenamiento y manipulación en los laboratorios hasta su uso, afecte más negativamente a los medios de cultivo que a las puntas de pipeta (Nijs et al., 2009; Swain et al., 2012;). Esto podría provocar que medios que no eran tóxicos en origen se hayan alterado, dando resultados en TSE diferentes a los que debería obtenerse.

En nuestro trabajo no hemos ajustado por la experiencia previa de los laboratorios en TSE, pero el hecho de no haber observado relación entre el porcentaje de aciertos en un laboratorio con medios de cultivo y el porcentaje de aciertos con puntas de pipeta, sugiere, que las diferencias observadas se deben más al tipo de producto que a un factor de laboratorio. Este último hubiera hecho observar una relación entre resultados con puntas de pipeta y medios de cultivo al analizarlos por laboratorios.

Desde un punto de vista logístico, resulta más fácil la preparación y envío de puntas que de medios de cultivo. Lo que, añadido a los resultados obtenidos, hacen de éste el producto el ideal a utilizar en los CCExt de CT. No obstante, el uso de

puntas de pipeta impediría realizar al mismo tiempo CCExt de pH y osmolaridad como ocurre en otros programas CCExt de CT que envían medios de cultivo (AAB, 2017).

En conclusión, creemos que el uso de materiales como puntas de pipeta es más adecuado para programas de CCExt, pues permiten obtener resultados más comparables entre laboratorios. El envío de medios de cultivo debería estar más enfocado a CCExt de parámetros químicos como pH y osmolaridad que a test de CT.

Conclusiones

1. La evaluación mediante morfología clásica de embriones sin anomalías morfológicas se realiza de forma fiable en los laboratorios de embriología. Cuando se observan diferencias entre laboratorios de embriología y expertos, se deben a la presencia de vacuolas o multinucleación en embriones en estadio de células y a masas celulares internas de baja calidad en estadio de blastocisto.
2. Es necesario mejorar la formación de los embriólogos clínicos en la evaluación de estas características morfológicas al utilizar los métodos clásicos de evaluación embrionaria.
3. Deberían analizarse las políticas de transferencia embrionaria de los centros con el fin de lograr niveles más altos de acuerdo respecto al destino final del embrión cuando es evaluado con métodos morfológicos clásicos, especialmente cuando el destino final de los embriones debe ser crioconservar.
4. La variabilidad entre laboratorios en la determinación de los tiempos a los que ocurren los eventos embrionarios depende de la duración del evento.
5. Los equipos *time-lapse* han disminuido la variabilidad entre laboratorios en la decisión clínica tomada sobre embriones. Sin embargo, no han modificado la variabilidad existente entre laboratorios en la catalogación de embriones.
6. Es necesario estandarizar la interpretación de las alteraciones morfológicas embrionarias observadas en equipos *time-lapse* antes de incorporar estas características a los algoritmos diagnósticos que utilizan dichos equipos.

7. El tipo de fungible empleado en los controles de calidad externos de citotoxicidad del laboratorio de reproducción humana asistida influye en los resultados de dichos controles.

8. La utilización de puntas de pipeta como fungible en un programa de control de calidad externo de citotoxicidad es fácil logísticamente y permite obtener resultados válidos en dichos programas. Sin embargo, no permite analizar controles de calidad de otras características (pH u osmolaridad) que pueden afectar a la viabilidad de gametos y embriones.

Bibliografía

- Aguilar J, Álvarez C, Morancho-Zaragoza J, Prats-Gimenez R, Ramírez JP, Fernández-Pardo E, et al. Quality specifications for seminal parameters based on clinicians' opinions. *Scand J Clin Lab Invest*. 2008;68:68–76.
- Aguilar J, Motato Y, Escribá MJ, Ojeda M, Muñoz E, Meseguer M. The human first cell cycle: impact on implantation. *Reprod Biomed Online*. 2014;28:475-484.
- Ahlstrom A, Park H, Bergh C, Selleskog U, Lundin K. Conventional morphological assessment performs better than morphokinetics for prediction of live birth after day 2 transfer. *Reprod Biomed Online*. 2016;33:61-70.
- Ahlstrom A, Westin C, Reisner E, Wikland M, Hardarson T. Trophoctoderm morphology: an important parameter for predicting live birth after single blastocyst transfer. *Hum Reprod*. 2011;26:3289–3296.
- Al-Hasani S, Ozmen B, Koutlaki N, Schoepper B, Diedrich K, Schultze-Mosgau A. Three years of routine vitrification of human zygotes: is it still fair to advocate slow-rate freezing? *Reprod Biomed Online*. 2007;14:288-293.
- Alikani M y Cohen J. Patterns of cell fragmentation in the human embryo. *J Assist Reprod Genet*. 1995;12:28.
- Alikani M, Calderon G, Tomkin G, Garrisi J, Kokot M, Cohen J. Cleavage anomalies in early human embryos and survival after prolonged culture in vitro. *Hum Reprod*. 2000;15:2634-2643.

Alpha Scientists in Reproductive Medicine y ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. *Hum Reprod.* 2011;26:1270-1283.

Álvarez C, Castilla JA, Martínez L, Ramírez JP, Vergara F, Gaforio JJ. Biological variation of seminal parameters in healthy subjects. *Hum Reprod.* 2003;18:2082-2088.

Álvarez C, Castilla JA, Ramírez JP, Vergara F, Yoldi A, Fernández A, et al. External quality control program for semen analysis: Spanish experience. *J Assist Reprod Genet.* 2005;22:379-387.

American Association of Bioanalysts. Statistical Summaries (2013-2016): AAB [Internet]. [Consultado el 20 de diciembre de 2017]. Disponible en: <http://www.aab-pts.org>

Aparicio B, Cruz M, Meseguer M. Is morphokinetic analysis the answer? *Reprod Biomed Online.* 2013;27:654-663.

Aparicio-Ruiz B, Basile N, Pérez Albalá S, Bronet F, Remohí J, Meseguer M. Automatic time-lapse instrument is superior to single-point morphology observation for selecting viable embryos: retrospective study in oocyte donation. *Fertil Steril.* 2016;106:1379-1385.

Arce JC, Ziebe S, Lundin K, Janssens R, Helmggaard L, Sørensen P. Interobserver agreement and intraobserver reproducibility of embryo quality assessments. *Hum Reprod.* 2006;21:2141-2148.

Arday M, Calderón G, Cuadros J, Figueroa MJ, Herrero R, Moreno JM, et al.

Cuadernos de Embriología Clínica. Criterios ASEBIR de valoración morfológica de oocitos, embriones tempranos y blastocistos humanos. 2ª edición. Madrid: ASEBIR; 2008.

Arroyo G, Veiga A, Santaló J, Barri PN. Developmental prognosis for zygotes based

on pronuclear pattern: usefulness of pronuclear scoring. J Assist Reprod Genet. 2007;24:173-181.

ASEBIR. CONTROL DE CALIDAD PARA EL LABORATORIO DE

REPRODUCCIÓN: ASEBIR [Internet]. [Consultado el 20 de diciembre de 2017].

Disponible en: <http://controldecualidad.asebir.com/index.php>.

Athayde Wirka K, Chen AA, Conaghan J, Ivani K, Gvakharia M, Behr B, et al.

Atypical embryo phenotypes identified by time-lapse microscopy: high prevalence and association with embryo development. Fertil Steril. 2014;101:1637-1648.

Avery S y Brinsden PR. Making the best use of human embryos. J Assist Reprod

Genet. 1999;16:457-459.

Aydin S, Cinar O, Demir B, Korkmaz C, Ozdegirmenci O, Dilbaz S, et al. Is

pronuclear scoring a really good predictor for ICSI cycles? Gynecol Endocrinol. 2011;27:742-747.

Azzarello A, Hoest T, Mikkelsen AL. The impact of pronuclei morphology and dynamicity on live birth outcome after time-lapse culture. *Hum Reprod.* 2012;27:2649-2657.

Balaban B, Urman B. Effect of oocyte morphology on embryo development and implantation. *Reprod Biomed Online.* 2006;12:608-615.

Balaban B, Yakin K, Urman B, Isiklar A, Tesarik J. Pronuclear morphology predicts embryo development and chromosome constitution. *Reprod Biomed Online.* 2004;8:695-700.

Basile N, Morbeck D, García-Velasco J, Bronet F, Meseguer M. Type of culture media does not affect embryo kinetics: a time-lapse analysis of sibling oocytes. *Hum Reprod.* 2013;28:634-641.

Basile N, Vime P, Florensa M, Aparicio Ruiz B, García Velasco JA, Remohí J, et al. The use of morphokinetics as a predictor of implantation: a multicentric study to define and validate an algorithm for embryo selection. *Hum Reprod.* 2015;30:276-283.

Basile N, Vime P, Florensa M, Aparicio Ruiz B, García Velasco JA, Remohí J, et al. The use of morphokinetics as a predictor of implantation: a multicentric study to define and validate an algorithm for embryo selection. *Hum Reprod.* 2015;30:276-283.

Baxter Bendus AE, Mayer JF, Shipley SK, Catherino WH. Interobserver and intraobserver variation in day 3 embryo grading. *Fertil Steril.* 2006;86:1608-1615.

Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med.* 1990;20:337-40.

Bras M, Lens J, Piederiet M, Rijnders P, Verveld M. *IVF Lab. Laboratory Aspects of In Vitro Fertilization.* Amsterdam: Organon; 1996.

Brezinova J, Oborna I, Svobodova M, Fingerova H. Evaluation of day one embryo quality and IVF outcome--a comparison of two scoring systems. *Reprod Biol Endocrinol.* 2009;7:9.

Cai X, Pomeroy KO, Mattox JH. Application study of human sperm motility bioassay in IVF laboratory quality control. *Zhonghua Nan Ke Xue.* 2006;12:625-628.

Calafell R, Domingo J, Fernández C et al. Especificaciones de calidad analítica. *An Clin.* 2002;27:95-100.

Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Thornton S. Retrospective analysis of outcomes after IVF using an aneuploidy risk model derived from time-lapse imaging without PGS. *Reprod Biomed Online.* 2013;27:140-146.

Castilla JA, Mantilla A. *Cuadernos de Embriología Clínica. Indicadores de calidad del laboratorio de embriología: definición y especificaciones.* 1ª Edición. Madrid: ASEBIR; 2016

- Castilla JA, Moracho-Zaragoza J, Aguilar J, Prats-Giménez R, Gonzalvo MC, Fernández-Prado E, et al. Quality specifications for seminal parameters based on the state of the art. *Hum Reprod.* 2005;20:2573-2578.
- Castilla JA, Ruiz de Assin R, Gonzalvo MC, Clavero A, Ramírez JP, Vergara F, et al. External quality control for embryology laboratory. *Reprod Biomed Online.* 2010;20:68-74.
- Chamayou S, Patrizio P, Storaci G, Tomaselli V, Alecci C, Ragolia C, et al. The use of morphokinetic parameters to select all embryos with full capacity to implant. *J Assist Reprod Genet.* 2013;30:703-710.
- Chan PJ, Calinisan JH, Corselli JU, Patton WC, King A. Updating quality control assays in the assisted reproductive technologies laboratory with a cryopreserved hamster oocyte DNA cytogenotoxic assay. *J Assist Reprod Genet.* 2001;18:129-134.
- Chavez SL, Loewke KE, Han J, Moussavi F, Colls P, Munne S, et al. Dynamic blastomere behaviour reflects human embryo ploidy by the four-cell stage. *Nat Commun.* 2012;13:1251.
- Chen AA, Tan L, Suraj V, Reijo Pera R, Shen S. Biomarkers identified with time-lapse imaging: discovery, validation, and practical application. *Fertil Steril.* 2013;99:1035-1043.
- Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *Clin Epidemiol.* 1990;43:551-558.

Cimadomo D, Ubaldi F, Capalbo A, Maggiulli R, Scaria C, Romano S, et al. Failure mode and effects analysis of witnessing protocols for ensuring traceability during PGD/PGS cycles. *Reprod Biomed Online*. 2016;33:360-369.

Ciray HN, Aksoy T, Goktas C, Ozturk B, Bahceci M. Time-lapse evaluation of human embryo development in single versus sequential culture media-a sibling oocyte study. *J Assist Reprod Genet*. 2012;29:891-900.

Ciray HN, Campbell A, Agerholm IE, Aguilar J, Chamayou S, Esbert M, et al. Time-lapse User Group. Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group. *Hum Reprod*. 2014;29:2650-2660.

Ciray HN, Campbell A, Agerholm IE, Aguilar J, Chamayou S, Esbert M, et al. Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group. *Hum Reprod*. 2014;29:2650–2660.

Ciray HN, Karagenc L, Ulug U, Bener F, Bahceci M. Early cleavage morphology affects the quality and implantation potential of day 3 embryos. *Fertil Steril*. 2006;85:358-365.

Ciray HN, Karagenç L, Ulug U, Bener F, Bahceci M. Use of both early cleavage and day 2 mononucleation to predict embryos with high implantation potential in intracytoplasmic sperm injection cycles. *Fertil Steril*. 2005;84:1411-1416.

Cobo A, Coello A, Remohí J, Serrano J, De Los Santos JM, Meseguer M. Effect of oocyte vitrification on embryo quality: time-lapse analysis and morphokinetic evaluation. *Fertil Steril*. 2017;108:491-497.

Cobo A, Kuwayama M, Pérez S, Ruiz A, Pellicer A, Remohi J. Comparison of concomitant outcome achieved with fresh and cryopreserved donor oocyte vitrified by the Cryotop method. *Fertil Steril*. 2007;89:1657-1664.

Cobo A, Pérez S, De los Santos MJ, Zulategui J, Domingo J, Remohí J. Effect of different cryopreservation protocols on the metaphase II spindle in human oocytes. *Reprod Biomed Online*. 2008;17:350-359.

Conaghan J, Chen AA, Willman SP, Ivani K, Chenette PE, Boostanfar R, et al. Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertil Steril*. 2013;100:412-419.

Cruz M, Garrido N, Herrero J, Pérez-Cano I, Muñoz M, Meseguer M. Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality. *Reprod Biomed Online*. 2012;25:371-381.

Cuadros M, Andrés Criado L, Sanchez de Burgos M, Morales Morales M, Ricciarelli E, Gómez Palomares JL, et al. Análisis comparativo de las tasas de implantación de embriones mononucleados y embriones multinucleados en D+2. *Rev Asoc Est Biol Rep*. 2011;16:168.

Cummins JM, Breen TM, Harrison KL, Shaw JM, Wilson LM, Hennessey JF. A formula for scoring human embryo growth rates in in vitro fertilization: its value in predicting pregnancy and in comparison with visual estimates of embryo quality. *J In Vitro Fert Embryo Transf.* 1986;3:284-295.

Dal Canto M, Coticchio G, Mignini Renzini M, De Ponti E, Novara PV, Brambillasca F, et al. Cleavage kinetics analysis of human embryos predicts development to blastocyst and implantation. *Reprod Biomed Online.* 2012;25:474-480.

De Jonge CJ, Centola GC, Reed ML, Shabanowitz RB, Simon SD, Quinn P. Human sperm survival assay as a bioassay for the assisted reproductive technologies laboratory. *J Androl.* 2003;24:16-18.

De los Santos MJ, Arroyo G, Busquet A, Calderón G, Cuadros J, Hurtado de Mendoza MV, et al. A multicenter prospective study to assess the effect of early cleavage on embryo quality, implantation, and live-birth rate. *Fertil Steril.* 2014;101:981-987.

De los Santos MJ, Gámiz P, Albert C, Galán A, Vilorio T, Pérez S, et al. Reduced oxygen tension improves embryo quality but not clinical pregnancy rates: a randomized clinical study into ovum donation cycles. *Fertil Steril.* 2013;100:402-407.

De Placido G, Wilding M, Strina I, Alviggi E, Alviggi C, Mollo A, et al. High outcome predictability after IVF using a combined score for zygote and embryo morphology and growth rate. *Hum Reprod.* 2002;17:2402-2409.

De Sutter P, Dozortsev D, Qian C, Dhont M. Oocyte morphology does not correlate with fertilization rate and embryo quality after intracytoplasmic sperm injection. *Hum Reprod.* 1996;11:595-597.

de Vet HC, Terwee CB, Bouter LM. Current challenges in clinimetrics. *J Clin Epidemiol.* 2003;56:1137-1141.

de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59:1033-1039.

Desai N, Ploskonka S, Goodman LR, Austin C, Goldberg J, Falcone T. Analysis of embryo morphokinetics, multinucleation and cleavage anomalies using continuous time-lapse monitoring in blastocyst transfer cycles. *Reprod Biol Endocrinol.* 2014;20:12:54.

Diamond MP, Suraj V, Behnke EJ, Yang X, Angle MJ, Lambe-Steinmiller JC, et al. Using the Eeva Test™ adjunctively to traditional day 3 morphology is informative for consistent embryo assessment within a panel of embryologists with diverse experience. *J Assist Reprod Genet.* 2015;32:61-68.

Donabedian A. The definition of quality and approaches to its assessment. *Ann Arbor. Michigan: Health Administration Press; 1980.*

Donner A, Eliasziw M. Statistical implications of the choice between a dichotomous or continuous trait in studies of interobserver agreement. *Biometrics.* 1994;50:550–555.

- Dunn G. Statistical evaluation of measurement errors: design and analysis of reliability studies. 2ª edición. London, UK: Arnold; 2004.
- Ehrmeyer U, Laessig RH. Has compliance with CLIA requirements really improved quality in US clinical laboratories? *Chim Acta*. 2001;346:37-43.
- Elder KT, Kastrop P. Control de calidad en laboratorios de fertilización in vitro. *Reproducción Humana [Internet]* 2003; 3 Vol 1: 13-20. [Consultado el 20 de diciembre de 2017] Disponible en: http://www.flasef.org/textos/revista/2003/1/13_20.pdf
- Emiliani S, Fasano G, Vandamme B, Vannin AS, Verdoodt M, Biramane J, et al. Impact of the assessment of early cleavage in a single embryo transfer policy. *Reprod Biomed Online*. 2006;13:255-260.
- ESHRE Capri Workshop Group. Intrauterine insemination. *Hum Reprod Update*. 2009;15:265-277.
- ESHRE Guideline Group on Good Practice in IVF Labs, De los Santos MJ, Apter, S., Coticchio, G., Debrock, S., Lundin, K., et al. Revised guidelines for good practice in IVF laboratories (2015). *Hum Reprod*. 2016;31,685–686.
- ESHRE Special Interest Group of Embryology and Alpha Scientists in Reproductive Medicine. The Vienna consensus: report of an expert meeting on the development of ART laboratory performance indicators. *Reprod Biomed Online*. 2017;35:494-510.

European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE), Calhaz-Jorge C, de Geyter C, Kupka MS, de Mouzon J, Erb K, et al. Assisted reproductive technology in Europe, 2012: results generated from European registers by ESHRE. *Hum. Reprod.* 2016;31:1638-1652.

Fauque P, Audureau E, Leandri R, Delaroche L, Assouline S, Epelboin S, et al. Is the nuclear status of an embryo an independent factor to predict its ability to develop to term? *Fertil Steril.* 2013;99:1299-1304.

Federación internacional de química clínica. Recomendación aprobada sobre control de calidad en química clínica. Parte 5. Control de calidad externo. *Quim Clin.* 1983;2:87-93.

Feinstein AR, Cicchetti DV. High agreement but low kappa: I. the problems of two paradoxes. *J Clin Epidemiol.* 1990;43:543-549.

Fenwick J, Platteau P, Murdoch AP, Herbert M. Time from insemination to first cleavage predicts developmental competence of human preimplantation embryos in vitro. *Hum Reprod.* 2002;17:407-412.

Ferré M, Fuentes X. Control interno de la calidad. En: Fuentes X, Castiñeiras MJ, Queraltó JM. *Bioquímica clínica y patología molecular. Volumen I. 2ª edición.* Barcelona: Editorial Reverte; 1998.

Figueira RCS, Ferreira Braga DP, Semião-Francisco L, Madaschi C, Iaconelli A Jr.ç, Borges E Jr. Metaphase II human oocyte morphology: contributing factors and

effects on fertilization potential and embryo developmental ability in ICSI cycles. *Fertil Steril.* 2010;94:1115-1117.

Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas.* 1973;33:613-619

Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions.* 3ª edición. Hoboken, NJ: Wiley; 2003.

Fraser CG, Hyltoft-Peterson P, Ricós C, Haeckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. *Eur J Clin Chem Clin Biochem.* 1992;30:311-317.

Fraser CG. *Biological variation: from principles to practice.* 1ª edición. Washington, USA: AACC Press; 2001.

Gardner DK, Balaban B. Assessment of human embryo development using morphological criteria in an era of time-lapse, algorithms and 'OMICS': is looking good still important? *Mol. Hum Reprod.* 2016;22:704-718.

Gardner DK, Reed L, Linck D, Sheehan C, Lane M. Quality control in human in vitro fertilization. *Semin Reprod Med.* 2005;23:319-324.

Giménez A, Molina P, Ruíz J.J, Acosta F, López M, Jiménez M, et al. Aplicación del análisis modal de fallos y sus efectos a la fase preanalítica de un laboratorio clínico. *Rev Lab Clin.* 2010;4:161–170.

Giorgetti C, Terriou P, Auquier P, Hans E, Spach JL, Salzmann J, et al. Embryo score to predict implantation after in-vitro fertilization: based on 957 single embryo transfers. *Hum Reprod.* 1995;10:2427-2431.

Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Res Social Adm Pharm.* 2013;9:330-338.

Guerif F, Le Gouge A, Giraudeau B, Poindron J, Bidault R, Gasnier O, et al. Limited value of morphological assessment at days 1 and 2 to predict blastocyst development potential: a prospective study based on 4042 embryos. *Hum Reprod.* 2007;22:1973-1981.

Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol.* 2008;61:29-48.

Haeckel R. Evaluation methods in laboratory medicine. Weinheim, Germany: VCH; 1993.

Hardarson T, Hanson C, Sjogren A, Lundin K. Human embryos with unevenly sized blastomeres have lower pregnancy and implantation rates: indications for aneuploidy and multinucleation. *Hum Reprod.* 2001;16:313-318.

Hardarson T, Löfman C, Coull G, Sjögren A, Hamberger L, Edwards RG. Internalization of cellular fragments in a human embryo: Time-lapse recordings. *Reprod Biomed Online.* 2002;5:36-38.

- Hardarson T, Löfman C, Coull G, Sjögren A, Hamberger L, Edwards RG. Internalization of cellular fragments in a human embryo: Time-lapse recordings. *Reprod Biomed Online*. 2002;5:36-38.
- Harper J, Magli MC, Lundin K, Barratt CL, Brison D. When and how should new technology be introduced into the IVF laboratory? *Hum Reprod*. 2012;27:303-313
- Harris EK, Boyd JC. Statistical bases of reference values in laboratory medicine. New York: Marcel Dekker Inc; 1995
- Hashimoto S, Kato N, Saeki K, Morimoto Y. Selection of high-potential embryos by culture in poly(dimethylsiloxane) microwells and time-lapse imaging. *Fertil Steril*. 2012;97:332-337.
- Hernaez R. Reliability and agreement studies: a guide for clinical investigators. *Gut*. 2015;64:1018-1027.
- Herrero J, Tejera A, Albert C, Vidal C, de los Santos MJ, Meseguer M. A time to look back: analysis of morphokinetic characteristics of human embryo development. *Fertil Steril*. 2013;100:1602-1609.
- Hill GA, Freeman M, Bastias MC, Rogers BJ, Herbert CM, Osteen KG, et al. The influence of oocyte maturity and embryo quality on pregnancy rate in a program for in vitro fertilization-embryo transfer. *Fertil Steril*. 1989;52:801-806.

Hill MJ, Richter KS, Heitmann RJ, Graham JR, Tucker MJ, DeCherney AH, et al. Trophoctoderm grade predicts outcomes of single-blastocyst transfers. *Fertil Steril*. 2013;99:1283-1289.

Hlinka D, Kalatova B, Uhrinova I, Dolinksa S, Rutarova J, Rezacova J, et al. Time-lapse cleavage rating predicts human embryo viability. *Physiol Res*. 2012;61:513-525.

Hnida C, Engenheiro E, Ziebe S. Computer-controlled, multilevel, morphometric analysis of blastomere size as biomarker of fragmentation and multinuclearity in human embryos. *Hum Reprod*. 2004;19:288-293.

Hoehler FK. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol*. 2000;53:499-503.

Holte J, Berglund L, Milton K, Garello C, Gennarelli G, Revelli A, et al. Construction of an evidence-based integrated morphology cleavage embryo score for implantation potential of embryos scored and transferred on day 2 after oocyte retrieval. *Hum Reprod*. 2007;22:548-557.

Hossain A, Aryal S, Osuampke C, Phelps J. Human sperm bioassay for reprotoxicity testing in embryo culture media: some practical considerations in reducing the assay time. *Adv Urol*. 2010;2010:136898.

Hurtado de Mendoza MV, Cuadros J, de los Santos MJ, Figueroa MJ, Prados F, et al. Cuadernos de embriología clínica. Criterios ASEBIR de Valoración

Morfológica de Oocitos, Embriones Tempranos y Blastocistos Humanos. 3ª edición. Madrid; ASEBIR; 2015.

Iemmolo M, Simmons L, Matson P. The rapid detection of cytotoxicity using a modified human sperm survival assay. J Assist Reprod Genet. 2005;22:177-180.

International Union of Pure and Applied Chemistry (IUPAC). The international armonized protocol for the proficiency testing of (chemical) analytical laboratories. Pure Appl Chem. 1996;65:2123-2144.

Intra G, Alteri A, Corti L, Rabellotti E, Papaleo E, Restelli L, et al. Application of failure mode and effect analysis in an assisted reproduction technology laboratory. Reprod Biomed Online. 2016;33:132-139.

ISO 13528. Statistical methods for use in proficiency testing by interlaboratory comparisons. Geneva: International Organization for Standardization; 2015.

ISO 15189. Medical laboratories – Particular requirements for quality and competence. Geneva: International Organization for Standardization; 2012.

ISO 9001. Quality systems. Model for quality assurance in design, development, production installation and servicing. Geneva: International Organization for Standardization; 2015.

ISO 9004. Quality management and quality systems elements. Part I. Guidelines. Geneva: International Organization for Standardization; 2009.

James AN, Hennessy S, Reggio B, Wiemer K, Larsen F, Cohen J. The limited importance of pronuclear scoring of human zygotes. *Hum Reprod.* 2006;21:1599-1604.

Johansson M, Hardarson T, Lundin K. There is a cutoff limit in diameter between a blastomere and a small anucleate fragment. *J Assist Reprod Genet.* 2003;20:309-313.

Kaser DJ, Racowsky C. Clinical outcomes following selection of human preimplantation embryos with time-lapse monitoring: a systematic review. *Hum Reprod Update.* 2014;20:617-631.

Keck C, Fischer R, Baukloh V, Alper M. Staff management in the in vitro fertilization laboratory. *Fertil Steril.* 2005;84:1786-1788.

Kenny D, Fraser CG, Hyltoft-Petersen PH, Kallner A. Consensus agreement. *Scand J Clin Lab Invest.* 1999;59:585.

Kieslinger DC, De Gheselle S, Lambalk CB, De Sutter P, Kosteljik EH, Twisk JW, et al. Embryo selection using time-lapse analysis (Early Embryo Viability Assessment) in conjunction with standard morphology: a prospective two-center pilot study. *Hum Reprod.* 2016;31:2450-2457.

Kim JJ, Patton WC, Corselli J, Jacobson JD, King A, Chan PJ. Mouse embryonic stem cells for quality control testing in assisted reproductive technology programs. *J Reprod Med.* 2005;50:533-538.

Kirkegaard K, Agerholm IE, Ingerslev HJ. Time-lapse monitoring as a tool for clinical embryo assessment. *Hum Reprod.* 2012;27:1277-1285.

Kirkegaard K, Campbell A, Agerholm I, Bentin-Ley U, Gabrielsen A, Kirk J, et al. Limitations of a time-lapse blastocyst prediction model: a large multicentre outcome analysis. *Reprod Biomed Online.* 2014;29:156-8.

Kirkegaard K, Campbell A, Agerholm I, Bentin-Ley U, Gabrielsen A, Kirk J, et al. Limitations of a timelapse blastocyst prediction model: a large multicentre outcome analysis. *Reprod Biomed Online.* 2014;29:156-158.

Kirkegaard K, Hindkjaer JJ, Ingerslev HJ. Effect of oxygen concentration on human embryo development evaluated by time-lapse monitoring. *Fertil Steril.* 2013;99:738-744.

Kirkegaard K, Kesmodel US, Hindkjær JJ, Ingerslev HJ. Time-lapse parameters as predictors of blastocyst development and pregnancy outcome in embryos from good prognosis patients: a prospective cohort study. *Hum Reprod.* 2013;28:2643-2451.

Kohn LT, Corrigan JM, Donaldson MS. *To Err is Human: Building a Safer Health System.* National Academies Press; 2000.

Kop PA, van Wely M, Mol BW, de Melker AA, Janssens PM, Arends B, et al. Intrauterine insemination or intracervical insemination with cryopreserved donor sperm in the natural cycle: a cohort study. *Hum Reprod.* 2015;30:603-607.

Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud.* 2011;48:661-671.

Kraemer HC, Periyakoil VS, Noda A. Kappa coefficients in medical research. *Stat Med.* 2002;21:2109-129.

Krisher RL. The effect of oocyte quality on development. *J Anim Sci.* 2004;82 E-Suppl:E14-23.

Lagalla C, Barberi M, Orlando G, Sciajno R, Bonu MA, Borini A. A quantitative approach to blastocyst quality evaluation: morphometric analysis and related IVF outcomes *J Assist Reprod Genet.* 2015;32:705–712.

Lan KC, Huang FJ, Lin YC, Kung FT, Hsieh CH, Huang HW, et al. The predictive value of using a combined Z-score and day 3 embryo morphology score in the assessment of embryo survival on day 5. *Hum Reprod.* 2003;18:1299-1306.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.

Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. *J Clin Epidemiol.* 1996;49:431-434.

Larman MG, Minasi MG, Rienzi L, Gardner DK. Maintenance of the meiotic spindle during vitrification in human and mouse oocytes. *Reprod Biomed Online.* 2007;15:692-700.

Lemmen JG, Agerholm I, Ziebe S. Kinetic markers of human embryo quality using timelapse recordings of IVF/ICSI-fertilized oocytes. *Reprod Biomed Online*. 2008;17:385-391.

Lierman S, De Sutter P, Dhont M, Van der Elst J. Double-quality control reveals high-level toxicity in gloves used for operator protection in assisted reproductive technology. *Fertil Steril*. 2007;88:1266-1272.

Liu SY, Teng B, Fu J, Li X, Zheng Y, Sun XX. Obstetric and neonatal outcomes after transfer of vitrified early cleavage embryos. *Hum Reprod*. 2013;28:2093-2100.

Liu Y, Chapple V, Feenan K, Roberts P, Matosn P. Time-lapse deselection model for human day 3 in vitro fertilization embryos: the combination of qualitative and quantitative measures of embryo growth. *Fertil Steril*. 2016;105:656–662

Liu Y, Chapple V, Feenan K, Roberts P, Matson P. Clinical significance of intercellular contact at the four-cell stage of human embryos, and the use of abnormal cleavage patterns to identify embryos with low implantation potential: a time-lapse study. *Fertil Steril*. 2015;103:1485-1491.

Liu Y, Chapple V, Feenan K, Roberts P, Matson P. Time-lapse videography of human embryos: Using pronuclear fading rather than insemination in IVF and ICSI cycles removes inconsistencies in time to reach early cleavage milestones. *Reprod Biol*. 2015;15:122-125.

López-Regalado ML, Castilla JA. Fichas de indicadores de calidad Abril 2017: ASEBIR [Internet]. [Consultado el 20 de diciembre de 2017]. Disponible en:

<http://asebir.com/wp-content/uploads/2014/06/FICHAS-DE-INDICADORES-DE-CALIDAD-Abril-2017.pdf>.

Lott JA. Setting process control limits for enzyme tests in serum. *Scand J Lab Clin Invest.* 1999;59:527-530.

Loutradis D, Drakakis P, Kallianidis K, Milingos S, Dendrinis S, Michalas S. Oocyte morphology correlates with embryo quality and pregnancy rate after intracytoplasmic sperm injection. *Fertil Steril.* 1999;72:240-244.

Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol.* 2010;63:854–61.

Lundin K, Ahlstrom A. Quality control and standardization of embryo morphology scoring and viability markers. *Reprod Biomed Online.* 2015;31:459-471.

Lundin K, Bergh C, Hardarson T. Early embryo cleavage is a strong indicator of embryo quality in human IVF. *Hum Reprod.* 2001;16:2652-2657.

Magli, M.C., Jones, G.M., Lundin, K., van den Abbeel, E. Atlas of human embryology: from oocytes to preimplantation embryos. *Hum. Reprod. Suppl* 1:i1. doi: 10.1093/humrep/des229. [Internet]. [Consultado el 20 de diciembre de 2017]. Disponible en: <http://atlas.eshre.eu/>

- Martínez-Burgos M, Herrero L, Megías D, Salvanes R, Montoya MC, Cobo AC, et al. Vitrification versus slow freezing of oocytes: effects on morphologic appearance, meiotic spindle configuration, and DNA damage. *Fertil Steril*. 2011;95:374-377.
- Mastenbroek S, Van der Veen F, Aflatoonian A, Shapiro B, Bossuyt P, Repping S. Embryo selection in IVF. *Hum Reprod*. 2011;26:964-966.
- Matava MJ, Arciero RA, Baumgarten KM, Carey JL, DeBerardino TM, Hame SL, et al. Multirater agreement of the causes of anterior cruciate ligament reconstruction failure: a radiographic and video analysis of the MARS cohort. *Am J Sports Med*. 2015;43:309-310.
- Meriano J, Clark C, Cadesky K, Laskin CA. Binucleated and micronucleated blastomeres in embryos derived from human assisted reproduction cycles. *Reprod Biomed Online*. 2004;9:511-520.
- Meseguer M, Herrero J, Tejera A, Hilligsoe KM, Ramsing N, Remohí J. The use of morphokinetics as a predictor of embryo implantation. *Hum Reprod*. 2011;26:2658-2671.
- Meseguer M, Rubio I, Cruz M, Basile N, Marcos J, Requena A. Embryo incubation and selection in a time-lapse monitoring system improves pregnancy outcome compared with a standard incubator: a retrospective cohort study. *Fertil Steril*. 2012;98:1481-1489.
- Milewski R, Kuć P, Kuczyńska A, Stankiewicz B, Łukaszuk K, Kuczyński W. A predictive model for blastocyst formation based on morphokinetic parameters in

time-lapse monitoring of embryo development. *J Assist Reprod Genet.* 2015;32:571-579.

Miller J, Morgan K, McAlister A, Freeman MR. A comparison between the human sperm bioassay and the mouse embryo bioassay. *Fertil Steril.* 2001;76:104–105.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19:539–549.

Molder A, Drury S, Costen N, Hartshorne GM, Czanner S. Semiautomated analysis of embryoscope images: Using localized variance of image intensity to detect embryo developmental stages. *Cytometry A.* 2015;87:119–128.

Molina I, Zamora S, López-Regalado ML, Castilla JA, et al. *Gestión de la calidad pretécnica en el laboratorio de reproducción humana asistida.* 1ª edición. Madrid: ASEBIR; 2016.

Montag M, Köster M, van der Ven K, van der Ven H. Gamete competence assessment by polarizing optics in assisted reproduction. *Hum Reprod Update.* 2011;17:654-666.

Mortimer, S.T., Mortimer, D., 2015. *Quality and Risk Management in the IVF Laboratory.* Cambridge. UK: Cambridge University Press; 2015.

Mulsant BH, Kastango KB, Rosen J, Stone RA, Mazumdar S, Pollock BG. Interrater reliability in clinical trials of depressive disorders. *Am J Psychiatry* 2002;159:1598-1600.

Munné S, Magli C, Adler A, Wright G, de Boer K, Mortimer D, et al. Treatment-related chromosome abnormalities in human embryos. *Hum Reprod.* 1997;12:780-784.

Nabitz U, Klazinga N, Walburg J. The EFQM excellence model: European and Dutch experiences with the EFQM approach in health care. *Int J Qual Health Care.* 2000;12:191-201.

Narayanan S. The preanalytic phase. An important component of laboratory medicine. *Am J Clin Pathol.* 2000;113:429-452.

Neuber E, Rinaudo P, Trimarchi JR, Sakkas D. Sequential assessment of individually cultured human embryos as an indicator of subsequent good quality blastocyst development. *Hum Reprod.* 2003;18:1307-1312.

Nijs M, Franssen K, Cox A, Wissmann D, Ruis H, Ombelet W. Reprotoxicity of intrauterine insemination and in vitro fertilization-embryo transfer disposables and products: a 4-year survey. *Fertil Steril.* 2009;92:527-535.

Organización Mundial de la Salud. WHO laboratory manual for the Examination and processing of human semen. 5ª edición. Geneva: OMS; 2010.

Ozgun K, Bulut H, Berkkanoglu M, Humaidan P, Coetzee K. Frozen embryo transfer can be performed in the cycle immediately following the freeze-all cycle. *J Assist Reprod Genet.* 2017. In press.

Paternot G, Devroe J, Debrock S, D'Hooghe TM, Spiessens C. Intra- and inter observer analysis in the morphological assessment of early-stage embryos. *Reprod Biol Endocrinol.* 2009;7:105.

Paternot G, Wetzels, AM, Thonon F, Vansteenbrugge A, Willemen D, Devroe J, et al. Intra- and interobserver analysis in the morphological assessment of early stage embryos during an IVF procedure: a multicentre study. *Reprod Biol Endocrinol.* 2011;9:127.

Payne D, Flaherty SP, Barry MF, Matthews CD. Preliminary observations on polar body extrusion and pronuclear formation in human oocytes using time-lapse video cinematography. *Hum Reprod.* 1997;12:532-541.

Petersen BM, Boel M, Montag M, Gardner DK. Development of a generally applicable morphokinetic algorithm capable of predicting the implantation potential of embryos transferred on day 3. *Hum Reprod.* 2016;31:2231-2244.

Petersen PH, Brandslund I, Stahl M, Defineolivarius N, Borch K, Ohnsen J. Evaluation of systematic and random factors in measurements of fasting plasma glucose as the basis for analytical quality specifications in the diagnosis of diabetes. Impact of the new WHO and ADA recommendations on diagnosis of diabetes mellitus. *Scand J Clin Lab Invest.* 2001;61:191-204.

Polit DF, Beck CT. Nursing research: generating and assessing evidence for nursing practice. 8ª edición. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.

Practice Committee of the American Society for Reproductive Medicine, Practice Committee of the Society for Assisted Reproductive Technology, Practice Committee of the Society of Reproductive Biology and Technology. Revised minimum standards for practices offering assisted reproductive technologies: a committee opinion. Fertil Steril. 2014;102:682-686.

Prados FJ, Debrock S, Lemmen JG, Agerholm I. The cleavage stage embryo. Hum Reprod. 2012;27 Suppl 1:i50-71.

Racowsky C, Combelles CM, Nureddin A, Pan Y, Finn A, Miles L, et al. Day 3 and day 5 morphological predictors of embryo viability. Reprod Biomed Online. 2003;6:323-331.

Racowsky C, Vernon M, Mayer J, Ball GD, Behr B, Pomeroy KO, et al. Standardization of grading embryo morphology. Fertil. Steril. 2010;94:1152-1153.

Rauch CA, Nichols JH. Laboratory accreditation and inspection. Clin Lab Med. 2007;27: 845-858.

Reader KL, Stanton JL, Juengel JL. The Role of Oocyte Organelles in Determining Developmental Competence. Biology. Biology (Basel). 2017;6.

- Richardson A, Brearley S, Ahitan S, Chamberlain S, Davey T, Zujovic L, et al. A clinically useful simplified blastocyst grading system. *Reprod Biomed Online*. 2015;31,523–530.
- Richter KS, Harris DC, Daneshmand ST, Shapiro BS. Quantitative grading of a human blastocyst: optimal inner cell mass size and shape. *Fertil Steril*. 2001;76:1157-1167.
- Rienzi L, Bariani F, Dalla M, Romano S, Scaria C, Maggiulli R, et al. Failure mode and effects analysis of witnessing protocols for ensuring traceability during IVF. *Reprod Biomed Online*. 2015;31:516-22.
- Rienzi L, Ubaldi F, Iacobelli M, Romano S, Minasi MG, Ferrero S, et al. Significance of morphological attributes of the early embryo. *Reprod Biomed Online*. 2005;10:669-681.
- Rienzi L, Vajta G, Ubaldi F. Predictive value of oocyte morphology in human IVF: a systematic review of the literature. *Hum Reprod Update*. 2011;17:34-45.
- Rijnders PM, Jansen CA. The predictive value of day 3 embryo morphology regarding blastocyst formation, pregnancy and implantation rate after day 5 transfer following in-vitro fertilization or intracytoplasmic sperm injection. *Hum Reprod*. 1998;13:2869-2873.
- Roque M. Freeze-all policy: is it time for that? *J Assist Reprod Genet*. 2015;32:171-176.

Rubio I, Kuhlmann R, Agerholm I, Kirk J, Herrero J, Escriba MJ, et al. Limited implantation success of direct-cleaved human zygotes: a time-lapse study. *Fertil Steril*. 2012;98:1458-1463.

Rubio I, Kuhlmann R, Agerholm I, Kirk J, Herrero J, Escribá MJ, et al. Limited implantation success of direct-cleaved human zygotes: a time-lapse study. *Fertil Steril*. 2012;98:1458-1463.

Ruiz de Assin R, Clavero A, Gonzalvo MC, Ramirez JP, Zamora S, Fernández A, et al. Comparison of methods to determine the assigned value in an external quality control programme for embryo evaluation. *Reprod Biomed Online*. 2009;19:824–829.

Ruiz de Assin R, Clavero A, Gonzalvo MC, Rosales A, Zamora S, Martínez L, et al. Reducing inter-observer variability in embryo evaluation by means of training courses. *J Assist Reprod Genet*. 2011;28:1129-33.

Saldeen P, Sundström P. Nuclear status of four-cell preembryos predicts implantation potential in in vitro fertilization treatment cycles. *Fertil Steril*. 2005;84:584-589.

Salumets A, Hyden-Granskog C, Makinen S, Suikkari AM, Tiitinen A, Tuuri T. Early cleavage predicts the viability of human embryos in elective single embryo transfer procedures. *Hum Reprod*. 2003;18:821-825.

- Salumets A, Hyden-Granskog C, Suikkari AM, Tiitinen A, Tuuri T. The predictive value of pronuclear morphology of zygotes in the assessment of human embryo quality. *Hum Reprod.* 2001;16:2177-2181.
- Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis WP, et al. Defining analytical performance specifications: consensus statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem Lab Med.* 2015;53:833–35.
- Santos de los, Gómez, DE, Castilla, JA, Ardoy, M. Estandarización de los Indicadores de Resultados en el Laboratorio de Reproducción Asistida. *Revista ASEBIR.* 2007;12:17-23.
- Sciacovelli L, Aita A, Plebani M. Extra-analytical quality indicators and laboratory performances. *Clin Biochem.* 2017;50:632-637.
- Scott L, Alvero R, Leondires M, Miller B. The morphology of human pronuclear embryos is positively related to blastocyst development and implantation. *Hum Reprod.* 2000;15:2394-2403.
- Shekhawat GS. Intrauterine insemination versus fallopian tube sperm perfusion in non-tubal infertility. *Med J Armed Forces India.* 2012;68:226-230.
- Shoukir Y, Campana A, Farley T, Sakkas D. Early cleavage of in-vitro fertilized human embryos to the 2-cell stage: a novel indicator of embryo quality and viability. *Hum Reprod.* 1997;12:1531-1536.

Shoukri MM. Measures of interobserver agreement. Boca Raton, FL: Chapman & Hall/CRC; 2004.

Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther.* 2005;85:257–268.

Steer CV, Mills CL, Tan SL, Campbell S, Edwards RG. The cumulative embryo score: a predictive embryo scoring technique to select the optimal number of embryos to transfer in an in-vitro fertilization and embryo transfer programme. *Hum Reprod.* 1992;7:117-119.

Stöckl D, Franzini C, Kratochvila J, Middle J, Ricós C, Siekmann L, et al. Analytical specifications of reference methods. Compilation and critical discussion. *Eur J Clin Chem Clin Biochem.* 1996;34:319-337.

Storr A, Venetis CA, Cooke S, Kilani S, Ledger W. Inter-observer and intra-observer agreement between embryologists during selection of a single Day 5 embryo for transfer: a multicenter study. *Hum Reprod.* 2017;32:307-314.

Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4^a edición. Oxford, UK: Oxford University Press; 2008.

Streiner DL. Clinimetrics vs. psychometrics: an unnecessary distinction. *J Clin Epidemiol.* 2003;56:1142-1145.

- Sundström P, Saldeen P. Early embryo cleavage and day 2 mononucleation after intracytoplasmic sperm injection for predicting embryo implantation potential in single embryo transfer cycles. *Fertil Steril*. 2008;89:475-477.
- Sundvall L, Ingerslev HJ, Knudsen UB, Kirkegaard K Inter- and intraobserver variability of time-lapse annotations. *Hum Reprod*. 2013;28:3215-21.
- Swain JE, Cabrera L, Xu X, Smith GD. Microdrop preparation factors influence culture-media osmolality, which can impair mouse embryo preimplantation development. *Reprod Biomed Online*. 2012;24:142–147
- Szklo M, Nieto FJ. *Epidemiology beyond the basics*. 3ª edición. Sudbury, MA: Jones and Bartlett; 2014.
- Tejera A, Herrero J, de Los Santos MJ, Garrido N, Ramsing N, Meseguer M. Oxygen consumption is a quality marker for human oocyte competence conditioned by ovarian stimulation regimens. *Fertil Steril*. 2011;96:618-623.
- Ten J, Mendiola J, Vioque J, de Juan J, Bernabeu R. Donor oocyte dysmorphisms and their influence on fertilization and embryo quality. *Reprod Biomed Online*. 2007; 14:40-48.
- Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42.

Tesarik J, Greco E. The probability of abnormal preimplantation development can be predicted by a single static observation on pronuclear stage morphology. Hum Reprod. 1999;14:1318-1323.

Tesarik J, Junca AM, Hazout A, Aubriot FX, Nathan C, Cohen-Bacrie P, et al. Embryos with high implantation potential after intracytoplasmic sperm injection can be recognized by a simple, non-invasive examination of pronuclear morphology. Hum Reprod. 2000;15:1396-1399.

The Alpha consensus meeting on cryopreservation key performance indicators and benchmarks: proceedings of an expert meeting. Reprod Biomed Online. 2012;25:146-167.

Tworek J, Nayar R, Savaloja L, Tabbara S, Thomas N, Winkler B, et al. General quality practices in gynecologic cytopathology: findings from the College of American Pathologists Gynecologic Cytopathology Quality Consensus Conference working group 3. Arch Pathol Lab Med. 2013;137:190-198.

UNE 179007. Health services. Systems of quality management for assisted reproduction laboratories. 2013.

Van Blerkom J, Davis P, Alexander S. Differential mitochondrial distribution in human pronuclear embryos leads to disproportionate inheritance between blastomeres: relationship to microtubular organization, ATP content and competence. Hum Reprod. 2000;15:2621-2633.

Van den Abbeel E, Balaban B, Ziebe S, Lundin K, Cuesta MJ, Klein BM, et al. Association between blastocyst morphology and outcome of single-blastocyst transfer. *Reprod Biomed Online*. 2013;27:353–361.

Van Royen E, Mangelschots K, De Neubourg D, Valkenburg M, Van de Meerssche M, Ryckaert G, et al. Characterization of a top quality embryo, a step towards single-embryo transfer. *Hum Reprod*. 1999;14:2345-2349.

Vassault A, Grafmeyer D, de Graeve J, Cohen R, Beaudonnet A and Bienvenu J. Analyses de biologie médicale: spécification et norms d'acceptabilité à l'usage de la validation de techniques. *Ann Biol Clin*. 1999;57:685–695.

VerMilyea MD, Tan L, Anthony JT, Conaghan J, Ivani, K, Gvakharia M, et al. Computer-automated time-lapse analysis results correlate with embryo implantation and clinical pregnancy: a blinded, multi-centre study. *Reprod Biomed Online*. 2014;29:729-736.

Wittemer C, Bettahar-Lebugle K, Ohl J, Rongières C, Nisand I, Gerlinger P. Zygote evaluation: an efficient tool for embryo selection. *Hum Reprod*. 2000;15:2591-2597.

Wong CC, Loewke KE, Bossert NL, Behr B, De Jonge CJ, Baer TM, et al. Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nat Biotechnol*. 2010;28:1115-1121.

Wongpakaran, N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol.* 2013;13:61.

Young DS. *Effects of preanalytical variables on clinical laboratory test.* 2ª edición. Washington: AACC Press; 1997.

Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, et al. The International Glossary on Infertility and Fertility Care. *Hum Reprod.* 2017;32:1786-1801.

Zhou XH. Correcting for verification bias in studies of a diagnostic test's accuracy. *Stat Methods Med Res.* 1998;7:337–353.

Ziebe S, Petersen K, Lindenberg S, Andersen AG, Gabrielsen A, Andersen AN. Embryo morphology or cleavage stage: how to select the best embryos for transfer after in-vitro fertilization. *Hum Reprod.* 1997;12:1545-1549.

Zollner U, Zollner KP, Hartl G, Dietl J, Steck T. The use of a detailed zygote score after IVF/ICSI to obtain good quality blastocysts: the German experience. *Hum Reprod.* 2002;17:1327-1333.

