Statistics: estadística

Esta es la parte del protocolo que más quebraderos nos da a los traductores. Empezaremos por tranquilizar al lector diciendo que no todos los traductores entienden siempre todo lo que traducen sobre estadística. Esta es una ley básica. Tampoco es este el lugar para recoger un glosario de estadística, por lo que el lector sólo encontrará aquí unos pocos términos que, en nuestra opinión, no están perfectamente recogidos en las demás referencias de consulta que facilitamos.

Poco podemos añadir a la clarísima exposición que el lector encontrará en la Wikipedia[108] y en este artículo.[109] Para afrontar la traducción de la parte estadística con buen humor, proponemos al lector que consulte esta página[110] sobre el aspecto más informal de la prueba exacta de Fisher. Y recomendamos encarecidamente la bitácora del Dr. Deng[111] sobre bioestadística y ensayos clínicos.

Los términos específicos de la estadística más habituales están estupendamente traducidos, y muchas veces explicados, en el GEC y en el *Diccionario crítico de dudas* de Navarro. También podemos contar con la ayuda del glosario multilingüe de terminología estadística del International Statistics Institute, que puede consultarse en Internet.[112]

A continuación ofrecemos algunas reflexiones sobre un conjunto de términos que, de manera subjetiva, hemos seleccionado por su interés para el traductor de un protocolo.

ADMINISTRATIVE CENSORING[113]	**OBSERVACIÓN INCOMPLETA POR CONCLUSIÓN DEL ESTUDIO** Hemos incluido este lema para advertir al traductor del riesgo de traducir de forma rápida e inopinada, aquí y en cualquier otra ocasión del protocolo: no es, desde luego, «censura administrativa», pero puede que tampoco sirva la traducción que hemos propuesto, porque el traductor deberá optar por una opción que recoja las peculiaridades administrativas y legislativas que rodeen al caso concreto.
ALLOCATION	**ASIGNACIÓN** En nuestro contexto, rara vez procede utilizar sinónimos de «asignación» (reparto, repartición, distribución, prorrateo).

[108] http://en.wikipedia.org/wiki/Medical_statistics (última comprobación: 14 de diciembre de 2011).
[109] http://lbe.uab.es/vm/sp/old/docs/metodologia/ensayo-clinico.pdf (última comprobación: 14 de diciembre de 2011).
[110] http://en.wikipedia.org/wiki/Lady_tasting_tea (última comprobación: 14 de diciembre de 2011).
[111] http://onbiostatistics.blogspot.com/ (última comprobación: 5 de enero de 2012).
[112] http://isi.cbs.nl/glossary/ (última comprobación: 27 de diciembre de 2011).
[113] Recordamos al lector que en la Wikipedia hay muchas páginas de calidad sobre estadística. Véase http://en.wikipedia.org/wiki/Censoring_(statistics) (última comprobación: 28 de diciembre de 2011).

ALL-PATIENTS-AS-TREATED (APaT) (POPULATION)

GRUPO DE TODOS LOS PACIENTES TRATADOS

Para el análisis estadístico es fundamental agrupar a los pacientes que se encuentren exactamente en unas mismas circunstancias determinadas para el análisis de sus datos. Por ejemplo, es imperativo separar a quienes fueron reclutados pero no llegaron a la aleatorización, por las razones que fueran; separar a quienes fueron aleatorizados y no llegaron a recibir el tratamiento de los que sí lo recibieron; a los que fueron reclutados y aleatorizados, y llegaron a recibir el tratamiento en una sola ocasión, de los que lo recibieron todas las veces previstas.

Para los análisis de la seguridad y de la eficacia de los datos puede optarse por alguno de los grupos antedichos (por ejemplo, *all randomized patients who received at least one dose of IV study therapy*). Esto depende de criterios estadísticos y es un refinamiento que tiene visos de irse complicando en el futuro, a medida que aumenta la complejidad de los protocolos. En este artículo de la Wikipedia hay algunas definiciones muy claras e interesantes.[114]

La traducción de este lema, que es lo que nos ocupa, dependerá de la definición de la población correspondiente, que por ahora casi siempre recoge el propio protocolo. Con criterios de simplicidad y economía terminológica, daremos con la traducción adecuada. Los revisores, acostumbrados a lidiar a menudo con este problema, suelen tener criterios muy fijos al respecto, por lo que recomendamos negociar con ellos. En el contexto de la estadística de los protocolos nos atrevemos a recomendar «grupo» para traducir *population* con fines de claridad, aunque los epidemiólogos pueden rebatir nuestra propuesta. Asimismo, aconsejamos al traductor que considere la posibilidad de prescindir de la traducción de la preposición *as,* y que demuestre su dominio del español en la elección de las perífrasis que se verá obligado a utilizar.

AS-OBSERVED

TAL Y COMO SE OBTUVIERON, TAL Y COMO SE OBSERVARON

Normalmente esta expresión califica al término *data* o a algún tipo de dato concreto, y se refiere al análisis de la respuestas observadas en los pacientes, tanto si siguen en el ensayo como si no. Algunos autores opinan que significa «en el orden en que se obtuvieron». El traductor elegirá en función del contexto y exhibirá su dominio del español en la elección de las perífrasis que se verá obligado a utilizar.

AS-RANDOMIZED

TAL Y COMO SE ALEATORIZARON

Según el GEC, la traducción correcta de *as-randomized population* es «población analizada por orden de aleatorización»; añade que también puede verse traducido como «población (analizada) según fue aleatorizada». En el análisis *as randomized* se considera a cada sujeto tratado con la intervención que le fue asignada aleatoriamente, tanto si fue la que realmente recibió como si no. Este análisis se contrapone al análisis *as treated* (véase), en el cual se considera a cada sujeto tratado con lo que

[114] http://en.wikipedia.org/wiki/Analysis_of_clinical_trials (última comprobación: 5 de enero de 2012).

	recibió en realidad, tanto si coincide con lo que le tocó en el sorteo como si no. El traductor elegirá en función del contexto y exhibirá su dominio del español en la elección de las perífrasis que se verá obligado a utilizar.
AS-TREATED	**TAL Y COMO SE TRATARON** Algunos autores opinan que significa «en el orden en que se trataron». El traductor elegirá en función del contexto y exhibirá su dominio del español en la elección de las perífrasis que se verá obligado a utilizar.
BASELINE OBSERVATION CARRIED FORWARD (BOCF)	**IMPUTACIÓN DE LA OBSERVACIÓN BASAL (BOCF)** Es una de las técnicas que se utilizan para suplir la ausencia de determinados datos. Estas técnicas cuentan con una creciente oposición entre los estadísticos porque, como puede comprenderse, propician la introducción de sesgos. Como en LOCF, WOCF y MVTF (véanse), se emplea el acrónimo en inglés; todos se recogen perfectamente explicados en esta bitácora.[115]
BIAS	**SESGO,** ERROR SISTEMÁTICO Aunque hay una definición puramente estadística, «sesgo: diferencia entre el valor esperado de un estimador y el verdadero valor del parámetro», en nuestro contexto es más frecuente verlo relacionado con la posibilidad de incurrir en errores sistemáticos y no intencionados en la recopilación de datos. La entrada del Diccionario de la RANM es magnífica. En el GEC se encuentran diversos tipos de *bias.*
CARRY, TO	**IMPUTAR** Véase *impute, to.*
CATEGORICAL DATA	**DATOS CUALITATIVOS** Aunque propone los sinónimos «datos nominales» y «datos categóricos», el GEC advierte de que, según Meinert, «en investigación clínica no son datos categóricos (contundentes, decisivos, indiscutibles) sino datos que representan una cualidad, característica o condición, como el sexo, la raza o el estado civil de una persona, que no son susceptibles de operaciones aritméticas ordinarias». (Véase *qualitative data.*)
CENSOR, TO	**CENSURAR,** LIMITAR En nuestro contexto, se refiere al hecho de «no poder observar algo por algún motivo» (*to prevent something being observed;* Day, 2007). En estadística, «censura» y «censurar» no guardan ninguna relación con el significado que tienen en el uso habitual del lenguaje.
CENSORING	**CENSURA** En estadística y en investigación médica, el término «censura» se refiere a las medidas que pueden adoptarse cuando: *a)* sólo se conoce parcialmente el valor de una medición o de una observación, o *b)* se observa

[115] http://onbiostatistics.blogspot.com/2010/08/locf-bocf-wocf-and-mvtf.html (última comprobación: 5 de enero de 2012).

	un valor que está fuera del intervalo del instrumento de medición. No debe confundirse con *truncation* ni con *rounding* (véanse). En estadística, «censura» y «censurar» no guardan ninguna relación con el significado que tienen en el uso habitual del lenguaje. El traductor que precise más aclaraciones las encontrará en la Wikipedia.[116]
CLEAN DATA	**DATOS REVISADOS,** DATOS CORREGIDOS, DATOS DEFINITIVOS A pesar de ser un término fácil y muy intuitivo, en estadística no suele utilizarse «datos limpios».
FUTILITY	**INUTILIDAD,** FUTILIDAD El GEC propone también la traducción por «intrascendencia», que nos parece muy adecuada para otros contextos. En el nuestro, la traducción más comprensible es la que aquí proponemos, «inutilidad», que sirve para relajar un poco el elevado tono técnico de las disertaciones estadísticas.
HYPOTHESIS	**HIPÓTESIS** En nuestro contexto, «hipótesis» no es «suposición de algo posible o imposible para sacar de ello una consecuencia» (Diccionario de la RAE), sino la expresión de los objetivos del estudio (o de algún otro elemento) en lenguaje matemático. El lector puede consultar en Internet[117] un glosario estupendo para entender todas las definiciones relacionadas con las hipótesis.
IMPUTATION	**IMPUTACIÓN** Se refiere a las técnicas que se utilizan para completar los datos que faltan. (Véase *impute, to.*)
IMPUTE, TO	**IMPUTAR** El GEC nos recuerda que, lejos de sus implicaciones legales, «en investigación clínica, significa completar los datos que faltan con valores presuntamente válidos obtenidos mediante procedimientos diversos, estimaciones, cálculos basándose en los datos disponibles, etc.; el ejemplo más frecuente es LOCF» (véase *LOCF*).
INTENTION-TO-TREAT	**(POR) INTENCIÓN DE TRATAR** Aunque esta traducción está más que consagrada, queremos advertir al traductor que poco tiene que ver con las «intenciones» de nadie, sustantivo que puede tener muchas connotaciones en español y que, en realidad, se refiere a una previsión inicial del ensayo: el grupo de sujetos que recibirán el tratamiento. Puede diferir del grupo que realmente lo reciba (por abandonos, fallecimientos, errores, etc.) y la diferencia afecta al análisis estadístico. *Intention-to-treat* actúa como adjetivo de *analysis* y de *population,* y en esos casos utilizamos la preposición «por». La

[116] http://en.wikipedia.org/wiki/Censoring_(statistics) (última comprobación: 5 de enero de 2012).
[117] http://www.stats.gla.ac.uk/steps/glossary/hypothesis_testing.html (última comprobación: 5 de enero de 2012).

	«población por intención de tratar» suele estar formada, como indica el GEC, por «todos los sujetos asignados de forma aleatoria a cada grupo de tratamiento (es decir, todos los sujetos inicialmente aleatorizados *[all randomized subjects]* para recibir un tratamiento específico), con independencia de que finalmente hayan recibido el tratamiento en cuestión de forma íntegra o parcial, de que no lo hayan recibido en absoluto por el motivo que fuere o de que hayan cambiado de tratamiento sobre la marcha».
LAST OBSERVATION CARRIED FORWARD (LOCF)	**IMPUTACIÓN DE LA ÚLTIMA OBSERVACIÓN REALIZADA** El GEC recoge otras buenas posibilidades de traducción (que desaconsejamos en nuestro afán unificador). También se aclara que *LOCF* «es un método de análisis que se utiliza cuando faltan datos debido a que el paciente abandonó el estudio. Se toman los últimos resultados disponibles (por ejemplo, los de la consulta 5, si fue la última a la que acudió el paciente) en reemplazo de los de la consulta final (que son los que faltan)». Esta técnica cuenta con una creciente oposición entre los estadísticos porque, como puede comprenderse, propicia la introducción de sesgos. Como en BOCF, WOCF y MVTF (véanse), se emplea el acrónimo en inglés; todos se recogen perfectamente explicados en esta bitácora.[118]
MATCH	**PAREJA,** PAR Se refiere solo a parejas *(pairs)* compatibles, con arreglo a determinados criterios que se citan en el protocolo.
MATCH, TO	**EMPAREJAR,** APAREAR Como se afirma en el GEC, «*to match* es hacer que el grupo experimental *(test group)* sea comparable al de referencia *(control group)* con respecto a una o más características o variables». Muy específica de los protocolos es la expresión *pair matching,* que solo podemos traducir por «emparejar», pero que aporta el matiz de que cada miembro de la pareja pertenece a un grupo distinto, el experimental (los «casos») y el de referencia (los «controles» o «testigos»), por ejemplo. En nuestro contexto, *matched* es sinónimo de *paired* (véase).
MATCHED	**EMPAREJADO,** PAREADO, APAREADO Fuera del apartado de estadística, pero en el contexto de los protocolos, puede significar «equivalente» (con un sentido de «compatible»), y en algunas ocasiones puede que no sea necesario traducirlo.
MISSING	**OMITIDO,** *FALTANTE, AUSENTE* Queremos advertir al traductor contra la inercia de traducir *missing* por «perdido», como se hace muchas veces en otros contextos. Todo lo que hay que saber sobre datos omitidos está en inglés en http://bit.ly/yACENt.

[118] http://onbiostatistics.blogspot.com/2010/08/locf-bocf-wocf-and-mvtf.html (última comprobación: 5 de enero de 2012).

MISSING VALUE TREATED AS FAILURE (MVTF)	**IMPUTACIÓN DE VALORES OMITIDOS COMO SI FUERAN FRACASOS** Como en LOCF, BOCF y WOCF (véanse), se emplea el acrónimo en inglés; todos se recogen perfectamente explicados en esta bitácora.[119] En un ensayo determinado puede estar justificado considerar como pacientes que no han respondido («fracasos») a todos los que se hayan retirado en la fase inicial. Las técnicas de imputación estadística cuentan con una creciente oposición entre los estadísticos porque, como puede comprenderse, propician la introducción de sesgos.
MULTIFACTORIAL	**MULTIFACTORIAL** Esta traducción no supone ninguna dificultad, pero queremos advertir al traductor sobre la posibilidad de confundirse con «multivariante», que se explica en *multivariate* (véase).
MULTIVARIATE	**MULTIVARIANTE** Esta forma de traducir *multivariate* (en referencia a la relación con múltiples variables) está en pleno debate: Navarro piensa ahora así, pero en la segunda edición de su *Diccionario crítico de dudas* prefería «multifactorial». Nosotros nos inclinamos por «multivariante» o por una locución prepositiva del tipo de «con múltiples variables». La cacofonía no debe hacernos dudar: *multivariate analysis of variance* es «análisis multivariante de la varianza». No debe confundirse con *multifactorial* (véase).
PAIRED	**PAREADO,** EMPAREJADO En realidad quiere decir «por parejas» o, en un sentido mucho más amplio, «en grupos de dos». Pero habitualmente solo lo encontraremos calificando a una prueba estadística o a la forma de llevarla a cabo, por lo que «pareado» suele encajar.
PERSON AT RISK	**PERSONA EXPUESTA** Advertimos al traductor de que es más que probable que el revisor prefiera una de estas dos traducciones incorrectas: «persona de riesgo» y «persona con riesgo». Pero basta pararse a pensar para preferir nuestra opción, que es la del GEC.
PERSON-TIME	**AÑO(S)-PERSONA** Es sinónimo de *person-year* (véase).
PERSON-YEAR	**AÑO(S)-PERSONA** Desde el punto de vista matemático, es igual hablar de «años-persona» que de «personas-año». Es un concepto estadístico y epidemiológico que conviene tener claro, y que se explica perfectamente en este artículo de José Tapia.[120]

[119] http://onbiostatistics.blogspot.com/2010/08/locf-bocf-wocf-and-mvtf.html (última comprobación: 5 de enero de 2012).
[120] http://medtrad.org/panacea/IndiceGeneral/n17-18_tradyterm-TapiaP.pdf (última comprobación: 5 de enero de 2012).

POWER	**POTENCIA,** PODER Representa la probabilidad de rechazar la hipótesis nula cuando es realmente falsa. En los ensayos suele calcularse el número de pacientes necesarios (el «tamaño de la muestra») para dotar al análisis de una determinada potencia.
p-VALUE	**VALOR *p*** No es «valor de *p»,* y la *«p»* siempre va en cursiva. Una vez más la Wikipedia[121] acude en ayuda del traductor que desee comprender este concepto. No es superfluo recordar aquí la importancia de extremar el cuidado con los símbolos ≤ y ≥ que el traductor encontrará a menudo en este contexto; es obligatorio utilizarlos como tales y no con subrayados. La *p* representa la probabilidad de obtener un resultado al menos tan extremo como el que realmente se ha obtenido (valor del estadístico calculado), suponiendo que la hipótesis nula es cierta. La hipótesis nula queda rechazada si el valor *p* asociado al resultado observado es igual o menor que el nivel de significación establecido, convencionalmente 0,05 o 0,01 (5% o 1%), punto que se denomina «potencia del contraste».
QUALITATIVE DATA	**DATOS CUALITATIVOS** Aunque propone los sinónimos «datos nominales» y «datos categóricos», el GEC advierte de que, según Meinert, «en investigación clínica no son datos categóricos (contundentes, decisivos, indiscutibles) sino datos que representan una cualidad, característica o condición, como el sexo, la raza o el estado civil de una persona, que no son susceptibles de operaciones aritméticas ordinarias». (Véase *categorical data*.)
RANGE	**INTERVALO,** AMPLITUD Dudamos de llegar a tiempo para detener la marea de «rangos» que inunda nuestros textos especializados, tanto en la traducción de *rank* como en la de *range* (véase). Pero cualquiera de las dos propuestas que hacemos aquí es más precisa y más útil que la traducción acrítica por «rango» y que, no obstante, en ocasiones puede estar justificada.
RANK	**ORDEN** Dudamos de llegar a tiempo para detener la marea de «rangos» que inunda nuestros textos especializados, tanto en la traducción de *rank* como en la de *range* (véase). Aunque se ve con muchísima frecuencia y no podemos decir que sea incorrecto en todos los casos, aconsejamos al traductor que no utilice «rango» si el contexto permite utilizar «orden».
ROUNDING	**REDONDEO** Desde el punto de vista de la traducción este término no ofrece ningún problema, pero queremos advertir al traductor de que, en nuestro contexto, el redondeo se aleja de la práctica aproximativa y burda que se emplea para los precios de las cosas en algunas circunstancias, y se

[121] http://en.wikipedia.org/wiki/P-value (última comprobación: 5 de enero de 2012).

refiere a un más o menos sofisticado artificio matemático que se emplea para representar un valor complejo de forma más corta, más simple o más explícita.[122] En estadística, el redondeo y el truncamiento son formas de censura (véase *censoring*).

SIGNIFICANCE	**SIGNIFICACIÓN (ESTADÍSTICA),** IMPORTANCIA El Diccionario de la RANM y el GEC informan muy bien de lo imprescindible para manejar este concepto. El traductor deberá estar muy atento para no confundir *significance, significant* y *significative,* que en nuestro contexto pueden referirse a la estadística o a la importancia de las cosas. La significación estadística es decisiva en muchos pasos clave del análisis de los datos, y el traductor está obligado a dominar este concepto para que la información correspondiente cobre pleno sentido en español; sólo así podrá resolver posibles confusiones del redactor, que las hay. La significación estadística suele expresarse por el nivel de significación *p* (véase *significance level*). Con independencia del término que haya utilizado el redactor en inglés, el traductor deberá determinar si se refiere a la importancia o a la significación estadística de aquello a lo que califica, y optar por el término español que recoja esa diferencia.
SIGNIFICANCE LEVEL	**NIVEL DE SIGNIFICACIÓN** Es la forma en que suele expresarse la significación estadística, por medio del valor *p* (que suele adoptar, por consenso, un valor mayor o menor que el 5% o que el 1%). (Véase *p-value.*)
SIGNIFICANT	**IMPORTANTE** En español, «importante» y «significativo» pueden ser sinónimos, pero no todo lo importante *(significant)* es estadísticamente significativo *(significative).* Los originales no suelen inducir a la confusión, pero conviene mantener la alerta.
SIGNIFICATIVE	**SIGNIFICATIVO** En español, «importante» y «significativo» pueden ser sinónimos, pero no todo lo importante *(significant)* es estadísticamente significativo *(significative).*
SOURCE	**FUENTE, ORIGINAL** En nuestro contexto, *source* puede ser sustantivo («fuente») o adjetivo («original»). De ahí el frecuente error de denominar «datos fuente» a lo que son «datos originales». La intromisión de otros campos semánticos muy dominantes, en los que se habla de «códigos fuente», ha llevado a muchos traductores a no hacer esta diferencia.
STRATIFICATION	**ESTRATIFICACIÓN** Véase *stratified randomization.*

[122] http://en.wikipedia.org/wiki/Rounding (última comprobación: 5 de enero de 2012).

STRATIFIED RANDOMIZATION	**ALEATORIZACIÓN ESTRATIFICADA** GEC: «Es un tipo de asignación aleatoria en el que primero se clasifica a los pacientes en determinados estratos o categorías, según unos criterios pronósticos conocidos, y luego los pacientes de cada categoría son asignados de forma independiente a uno de los grupos de estudio (experimental o de referencia), mediante un procedimiento de aleatorización propio. El objetivo es que los grupos de estudio tengan aproximadamente el mismo número de pacientes en cada categoría pronóstica o de gravedad. En los ensayos multicéntricos existe la modalidad de considerar que cada uno de los centros es un estrato aparte con asignación aleatoria independiente».
TRUNCATION	**TRUNCAMIENTO** Se trata de otra técnica de censura de datos en estadística. Su traducción no supone ningún problema, pero el concepto matemático no es sencillo. Se explica perfectamente en la Wikipedia.[123]
WEIGH, TO	**PONDERAR** En nuestro contexto se refiere al hecho de «compensar», sobre todo cuando se obtienen valores demasiado extremos en una medición. Por ejemplo, si todos los ensayos de un análisis tuvieran exactamente la misma precisión, bastaría con calcular la media para conocer la magnitud de los diversos efectos. Pero si, como suele ocurrir, unos son más precisos que otros, podemos calcular una media ponderada, que asigna más importancia *(weight)* a los estudios que aportan más información.
WEIGHED	**PONDERADO** En este contexto no es difícil toparse con términos derivados de *weigh* y de *weight* (no deben confundirse: en nuestro contexto, *to weigh* es «ponderar» y «pesar», y *weight* es un sustantivo, «peso» (véase *to weigh*).
WORST OBSERVATION CARRIED FORWARD (WOCF)	**IMPUTACIÓN DE LA PEOR OBSERVACIÓN** Como en LOCF, BOCF y MVTF (véanse), se emplea el acrónimo en inglés; todos se recogen perfectamente explicados en esta bitácora.[124] Las técnicas de imputación estadística cuentan con una creciente oposición entre los estadísticos porque, como puede comprenderse, propician la introducción de sesgos.

[123] http://en.wikipedia.org/wiki/Truncation_(statistics) (última comprobación: 5 de enero de 2012).
[124] http://onbiostatistics.blogspot.com/2010/08/locf-bocf-wocf-and-mvtf.html (última comprobación: 5 de enero de 2012).