

Ingeniería de Software e Ingeniería del Conocimiento:

*Tendencias de Investigación e
Innovación Tecnológica
en Iberoamérica*



Editores:

**Raúl A. Aguilar Vera
Julio C. Díaz Mendoza
Gerzon E. Gómez Cruz
Edwin León Bojórquez**

ISBN: 978-607-707-096-2



UADY
FACULTAD DE
MATEMÁTICAS
"Luz, Ciencia y Verdad"

*Ingeniería de Software e
Ingeniería del Conocimiento:*

*Tendencias de Investigación e
Innovación Tecnológica
en Iberoamérica*

Editores:

Raúl A. Aguilar Vera
Julio C. Díaz Mendoza
Gerzon E. Gómez Cruz
Edwin León Bojórquez.

ISBN: 978-607-707-096-2

Alfaomega Grupo Editor

Datos catalogaficos

**Tendencias de Investigacion e innovacion
Tecnologica en Iberoamerica**

Alfaomega Grupo Editor, S. A. de C. V., México

ISBN: 978-607-707-096-2

**Tendencias de Investigación e Innovación
Tecnológica en Iberoamérica**

Derechos Reservados Alfaomega Grupo Editor, S. A. de C. V., México

Alfaomega Grupo Editor, noviembre de 2010

© 2010 Alfaomega Grupo Editor, S. A. de C. V., México

Pitágoras 1139, Col. Del Valle, 03100, México, D. F.

Miembro de la Camara Nacional de la Industria Editorial Mexicana

Registro No. 2317

Pag. Web: <http://www.alfaomega.com.mx>

E-mail: atencioncliente@alfaomega.com.mx

ISBN: 978-607-707-096-2

Derechos Reservados

Esta obra es propiedad intelectual de su autor y los derechos de publicación en lengua española han sido legalmente transferidos al editor. Prohibida su reproducción parcial o total por cualquier medio sin permiso por escrito del propietario de los derechos del copyright.

Nota importante:

La información contenida en esta obra tiene un fin exclusivamente didáctico, y por lo tanto, no está previsto su aprovechamiento a nivel profesional o industrial. Las indicaciones técnicas y programas incluidos, han sido elaborados con gran cuidado por el autor y reproducidos bajo estrictas normas de control. ALFAOMEGA GRUPO EDITOR, S.A., DE C.V. no será jurídicamente responsable por: errores u omisiones, daños y perjuicios que se pudieran atribuir al use de la información comprendida en este libro, ni por la utilización indebida que pudiera dársele.

Impreso en México. Printed in Mexico.

Empresas del grupo:

México: Alfaomega Grupo Editor. S. A. de C. V. - Pitágoras 1139, Col. Del Valle, México, D.F.- C.P. 03100, Tel. (52-55) 5089-7740 - Fax: (52-55)-5575-2420/2490. Sin costo: 01-800-020-4396 E-mail: atencionalcliente@alfaomega.com.mx

Colombia: Alfaomega Colombiana, S. A. - Carrera 15 No. 64 A 29 - PBX (57-1) 2100122, Bogotá, Colombia, Fax: (57-1)-6068648 E-mail: scliente@alfaomega.com.co

Chile: Alfaomega Grupo Editor, S. A. - General del Canto 370 - Providencia, Santiago, Chile Tel. (56-2) 235-4248 - Fax: (56-2)-235-5786 - E-mail: agechile@alfaomega.cl

Argentina: Alfaomega Grupo Editor Argentina, S. A. - Paraguay 1307 P.B. "11", Buenos Aires, Argentina, C.P. 1057 - Tel.: (54-11) - 4811-7183/8352, E-mail: ventas@alfaomegaeditor.com.ar

Índice

Pág.

Ingeniería de Requisitos

Proceso de Educación de Requisitos en Proyectos de Explotación de Información	01
---	----

Florencia Pollo, Paola Britos, Patricia Pesado y Ramón Garcia

Aplicación de SixSigma en los Procesos de Ingeniería de Requisitos: Fase de Entrenamiento	12
---	----

Nahur Meléndez, José Gallardo, Claudio Meneses y Nahur Meléndez

Diseño y Construcción Software

PFDR: Un Proceso de Fases para el Desarrollo de RIAs basado en UWE y ADV	24
--	----

Víctor Hernández, Luis Martínez, y Giner Alor

Sobre la Programación Extrema y la Gestión de la Calidad	36
--	----

Rodrigo Haussmann y Vianca Vega

La Arquitectura de Información vinculada al proceso de desarrollo de software	46
---	----

Anolandy Díaz y Yuniel Rodríguez

Integración de Patrones de Diseño y Aplicaciones Móviles en un Sistema de Gestión para el Control de Mantenciones de Placas Catódicas.....	56
--	----

Wilson Castillo y Oscar Sandoval

Aspect-Oriented Architectural Evaluation for Java-based Systems	74
---	----

Eunice Martínez, Ulises Juárez & Giner Alor

Pruebas del Software

Comparando las Técnicas de Verificación Todos los Usos y Cubrimiento de Sentencias	85
--	----

Diego Vallespir, Carmen Bogado, Silvana Moreno y Juliana Herbert

Teste Colaborativo de Software Público Brasileiro	96
---	----

Paulo Siqueira, Adalberto Crespo, Miguel Argollo, Celso Barros y Mario Jino

Mejora de Procesos

Experiencia en la Implantación de MoProSoft en una Empresa Escolar: caso AvanTI.....	109
--	-----

María Astorga, Brenda Flores, Gloria Chávez, Mónica Lam y Araceli Justo

Adoptando los Procesos de la Categoría de Operación de COMPETISOFT a través de una Guía Basada en Plantillas	119
--	-----

Miguel Morales, Guadalupe Ibarguengoitia, Francisco Pino y Mario Piattini

Modelos de Gestión de Servicios de TI en las Pequeñas y Medianas Empresas: Una Revisión Sistemática.....	131
--	-----

Gerzon Gómez, Angel Gomez y Sarita Domínguez

Procesos de Software de la 29110 Guiados por Historias de Usuario.....	142
--	-----

Sergio Cárdenas, Francisco Pino, Guadalupe Ibarguengoitia y Mario Piattini

Ontología para el ciclo de vida de los procesos de negocio implementados con servicios.....	151
---	-----

Andrea Delgado, Francisco Ruiz e Ignacio García

Enterprise Architecture Responsibilities and People Roles.....	163
--	-----

Igor Aguilar, José Carrillo y Edmundo Tovar

Experience Factory Infrastructure as a basis for Knowledge Management in a Software Process Improvement Program.....	174
--	-----

Brenda Flores & Oscar Rodríguez

Herramientas y Técnicas de Software

Análisis de Taxonomías de Herramientas CASE y su Asociación con los Procesos Relacionados al Desarrollo y Mantenimiento de Software	185
---	-----

Sandra Gastelum y Brenda Flores

Framework to Provide HighlyAutomated UNDO Capabilities on Software Systems	194
--	-----

Hernán Merlino, Oscar Dieste, Patricia Pesado y Ramón García

Bases de Datos y Minería de Datos

Obtención de Clientes Potenciales del Servicio Internet Banda Ancha en una Empresa de Telecomunicaciones de Ecuador, aplicando una metodología de Minería de Datos	206
--	-----

Fernando Uyaguari

Modelo Preliminar para Almacenar y Recuperar Métricas Software Obtenidas Mediante Minería de Datos	217
--	-----

Enrique Luna, Marco Villalobos, Edgar Taya, Carlos Martínez y José Torres

Análisis de Rendimiento de los Algoritmos EquipAsso y Mate-tree: Dos Algoritmos de Minería de Uso de la Web	228
---	-----

Ricardo Timarán

Modelo de conocimiento para el hallazgo de indicadores de gestión en una unidad académica, utilizando técnicas de descubrimiento en base de datos (Knowledge Discovery in Databases o KDD).....	240
---	-----

Mary Bernal y Rossana Timaure

Ingeniería de Procesos de Explotación de Información	253
--	-----

Florencia Pollo, Paola Britos, Patricia Pesado y Ramón García

Identification of Noisy Data in Databases by Means of a Clustering Process	264
--	-----

Horacio Kuna, Ramón García y Francisco Villatoro

Métricas e Ingeniería de Software Empírica

Medición de la Productividad de Proyectos de Software Desarrollados en Dos Empresas Ecuatorianas.....	275
---	-----

Lohana Lema, Manuel Olvera y Mónica Villavicencio

Ingeniería de Software Empírica. Aplicabilidad de Métodos de Síntesis Cuantitativa	287
<i>Enrique Fernández, Florencia Pollo, Hernan Amatriain, Oscar Dieste, Patricia Pesado y Ramón García</i>	
Aplicaciones Innovadoras de las TIC	
BPEL como servicio de intermediación en WS-CDL	299
<i>Isaac Machorro, Giner Alor y Jesús Cruz</i>	
Servicios de Localización, Georeferenciación, y Mensajería a través de la Computación Móvil	310
<i>Daniel Arenas, Edward Sandoval, Juan García, Martha Gómez y Claudia Cáceres</i>	
Guía Informática Digital para Personas con Necesidades Especiales	321
<i>Juan Ucán, Mayra Cabrera, José Peña</i>	
Aplicaciones en Informática Educativa	
Usabilidad del juego de rescate para niños con problemas del lenguaje	332
<i>Gerson Escobedo y Carlos Miranda</i>	
Análisis y Diseño de un Videojuego para la Enseñanza de las Matemáticas en Educación Básica.....	342
<i>Francisco Madera y Luis Basto, Edgar Cambranes, Rocío Uicab, Pilar Rosado</i>	
Análisis comparativo del desarrollo de componentes para plataformas educativas de libre distribución.....	352
<i>Víctor Herrera, Danice Cano y Humberto Centurión</i>	
Objetos de Aprendizaje en Términos de Servicios Web	363
<i>Jaime Muñoz, René Santaolaya, Edgar Calvillo, Ricardo Mendoza y Olivia Fragoso</i>	

Ingeniería de Software Empírica. Aplicabilidad de Métodos de Síntesis Cuantitativa

E. Fernández, M. Pollo, H. Amatriain, O. Dieste, P. Pesado, R. García-Martínez

Programa de Doctorado en Ciencias Informáticas. Facultad de Informática. UNLP
Escuela de Posgrado. Facultad Regional Buenos Aires. UTN
Grupo de Ingeniería de Software Experimental. Facultad de Informática. UPM
Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP - CICGrupo de
Investigación en Sistemas de Información. Departamento Desarrollo Productivo y Tecnológico.
UNLa.

enriquefernandez@educ.ar, rgarcia@unla.edu.ar

Resumen. La síntesis cuantitativa consiste en combinar los resultados de varios estudios experimentales con el objeto de generar nuevas piezas de conocimiento. Estas nuevas piezas de conocimientos serán más generales y fiables que los resultados obtenidos por los estudios individuales, ya que dichas piezas de conocimiento están sustentadas por una mayor cantidad de evidencia empírica. El objetivo del presente trabajo es determinar cuáles de los métodos de agregación conviene aplicar en el contexto experimental que hoy día presenta la Ingeniería de Software Experimental.

Palabras clave. Agregación de experimentos. Síntesis Cuantitativa. Método de Monte Carlo. Ingeniería de software Experimental.

1. Introducción

La síntesis cuantitativa [1] consiste en combinar los resultados de varios estudios experimentales con el objeto de generar nuevas piezas de conocimiento. Estas nuevas piezas de conocimientos serán más generales y fiables que los resultados obtenidos por los estudios individuales, ya que dichas piezas de conocimiento están sustentadas por una mayor cantidad de evidencia empírica. En Ingeniería del Software (SE), la síntesis cuantitativa, se ha popularizado en los últimos años desde que fue propuesta por Basilli [2] en 1996 y el primer trabajo de meta-Análisis que se conoce es el desarrollado por Miller [3] que logró combinar 4 experimentos en 1999. A semejanza de lo que se hace en medicina, en SE, se utilizan las Diferencia Medias Ponderadas (WMD) [4] como método de síntesis. Es de hacer notar que para que el método WMD pueda aplicarse de forma

fiable requiere que el conjunto de estudios a agregar cumplan ciertas restricciones, entre otras: contener un número mínimo de experimentos, homogéneos y que reportan todos los parámetros estadísticos necesarios (medias, varianzas o desvíos estándar y cantidad de sujetos experimentales). Estas restricciones limitan fuertemente su aplicabilidad en el actual contexto experimental de la SE, donde las replicaciones de experimentos son muy reducidas [5;6] y los estudios no proporcionan los parámetros estadísticos necesarios debido a problemas de reporte [7;8]. Este hecho provoca disfunciones en la síntesis. Por ejemplo, en [9; 10] si bien se identificó un conjunto de experimentos importante, no se llegó a agregar los resultados de los estudios identificados debido a la falta de estandarización de variables respuesta y la baja calidad de los reportes publicados; en [3] solo se pudieron agregar de cuatro estudios experimentales y solo en [11] se logró hacer una síntesis realmente valiosa combinando 15 estudios experimentales.

Si bien WMD es el método de síntesis cuantitativa más difundido y recomendado en la mayoría de las ciencias [12], no es el único que existe. Restringiéndonos únicamente a los métodos cuantitativos, en [4] se propone el método Conteo de Votos Estadístico (SVC) como una alternativa menos restrictiva al WMD, mientras que en [13] se propone el Response Ratio (RR) como método alternativo al WMD. Pero, a la fecha, no se conoce ningún caso de aplicación de estos métodos en SE.

Dado que en algunos trabajos [14; 15] se ha observado que el desempeño de los métodos de Meta-Análisis varía con la cantidad de experimentos incluidos, así como también con la cantidad de sujetos que estos experimentos incluyen, y que en algunos casos el tamaño de la varianza puede cambiar la fiabilidad de los métodos, el objetivo del presente trabajo es determinar cuál o cuáles de los métodos conviene aplicar en el contexto experimental que hoy día presenta la SE. Este trabajo se encuentra estructurado de la siguiente forma: la sección 2 describe en que consiste la síntesis cuantitativa de experimentos y se detallan las principales características de los métodos a evaluar; la sección 3 describe la metodología de análisis a aplicar; la sección 4 se presenta los resultados obtenidos en el proceso de simulación; en la sección 5 detallan las conclusiones obtenidas.

2. Estado de La cuestión

2.1. Síntesis Cuantitativa de estudios Experimentales

La síntesis cuantitativa de estudios experimentales, también llamada agregación de experimentos o Meta-Análisis, consiste en la integración de los resultados de un conjunto de experimentos, previamente identificados, que analizan el desempeño de un par de tratamientos predefinidos con el fin de dar una estimación cuantitativa sintética de todos los estudios disponibles [16]. Como el objeto de estudio de este tipo de trabajo son los estudios experimentales previamente desarrollados y analizados por sus autores, a este

tipo de estudio también se lo conoce con el nombre de “Meta-Análisis”, que significa “después del análisis” [17].

Si todos los estudios incluidos en el proceso de Meta-Análisis fueran igualmente precisos y utilizaran exactamente las mismas variables respuesta, bastaría con promediar los resultados de cada uno de ellos para obtener así una conclusión final [18]. Sin embargo, en la práctica no todos los estudios tienen la misma precisión, por ello cuando se los combine se debe asignar un mayor peso a los estudios que permiten obtener información más fiable. Esto se logra combinando los resultados mediante un promedio ponderado [19]. Por otra parte, para poder solucionar los problemas vinculados a la no uniformidad de las variables respuesta, los métodos Meta-Analíticos expresan sus resultados mediante un índice de “Tamaño de Efecto”, el cual es un estimador no escalar de la relación entre una exposición y un efecto [16] y es aplicable a cualquier medida de diferencia de los resultados de dos grupos.

El método de síntesis cuantitativa para variables continuas (las más utilizadas en SE) más utilizado es diferencias medias ponderadas (WMD) [4] (recomendado por organismos internacionales como Cochrane Collaboration [16]). No obstante existen otros métodos alternativos menos difundidos para el cálculo del tamaño de efecto, como son: el Response Ratio (RR) versión paramétrica propuesto por [13], el Vote Counting (VC) propuesto por [4] y el Response Ratio (RR) versión no paramétrica propuesto por [13]. Estos métodos se describen a continuación.

2.1.1 Diferencia Medias Ponderadas

Este método es conceptualmente sencillo: el estimador de efecto individual (representa la tasa de mejora de un tratamiento respecto del otro en cada experimentos) se estima como el cociente de las diferencias entre las medias y el desvío estándar conjunto:

$$g = \frac{Y^E - Y^C}{S_p} \quad \begin{array}{l} g \text{ es el tamaño de efecto} \\ Y^E \text{ y } Y^C \text{ son las medias de los tratamientos} \\ S_p \text{ es la varianza conjunta} \end{array} \quad (1)$$

$$S_p = \sqrt{\frac{(n^E - 1)(s^E)^2 + (n^C - 1)(s^C)^2}{n^E + n^C - 2}} \quad \begin{array}{l} S_p \text{ desvío estándar conjunto} \\ S^E \text{ y } S^C \text{ son el desvío estándar del grupo} \\ \text{experimental (E) y de control (C)} \\ n^E \text{ y } n^C \text{ son el número de sujetos del grupo} \\ \text{experimental (E) y de control (C)} \end{array} \quad (2)$$

Esta función (1), desarrollada por Glass [17], fue optimizada por Hedgges y Olkin [4] quienes incorporaron un factor de corrección que aumenta la fiabilidad cuando se trabaja con pocos estudios (3). Convirtiendo a la nueva función en el método de Meta-Análisis más difundido en la actualidad y el recomendado para ser utilizado en IS [2].

$$d = J(N - 2) \frac{Y^E - Y^C}{S_p} \quad \begin{array}{l} d \text{ es el Effect Size} \\ J(N - 2) = \text{factor de corrección} \\ Y^E \text{ y } Y^C \text{ son la media de los grupos experimental (E) y de} \\ \text{control (C)} \end{array} \quad (3)$$

S_p es el desvío estándar conjunto
 N es el número de sujetos conjunto ($n_E + n_C$)

$$J = 1 - \frac{3}{4N - 9} \quad \begin{array}{l} J \text{ es el factor de ajuste} \\ N \text{ es el número de sujetos conjunto } (n_E + n_C) \end{array} \quad (4)$$

Una vez estimado el tamaño de efecto para cada estudio, puede estimarse el *efecto global* (representa la tasa de mejora de un tratamiento respecto del otro a nivel general) el cual se calcula como una media ponderada de los estimadores de efecto de los estudios individuales [4]:

$$d^* = \frac{\sum d_i / \sigma_i^2(d)}{\sum 1 / \sigma_i^2(d)} \quad \begin{array}{l} d^* \text{ es el tamaño de efecto global} \\ \sum d_i / \sigma_i^2(d) \text{ es la suma de los efectos individuales} \\ \sum 1 / \sigma_i^2(d) \text{ es la suma de la inversa varianza} \end{array} \quad (5)$$

A la función (5) se la conoce también con el nombre de modelo de efectos fijo, porque asume que la variación entre los resultados de los experimentos se debe únicamente al error experimental. Existen perfeccionamientos de esta función para otro tipo de variaciones, tales como sesgos de publicación o heterogeneidad experimental. No obstante en este trabajo solo abordaremos el modelo de efecto fijo dado que el bajo número de experimentos típicamente incluidos en los meta análisis en IS hacen muy difícil estimar con precisión las variables (por ejemplo: la varianza entre estudios) requeridas por dichos perfeccionamientos (por ejemplo modelo de efectos aleatorios) [18].

Uno de los inconvenientes del tamaño de efecto estimado utilizando WMD (inconveniente que no posee, por ejemplo, el tamaño de efecto estimado por RR) es que su interpretación no es intuitiva: Esto es, no es inmediato saber cuánto de mejor es un tratamiento en relación a otro. En general se asume que un resultado igual a 0 implica un efecto nulo (los tratamientos se comportan de la misma forma), un resultado igual a 0.2 implica un tamaño de efecto pequeño (uno de los tratamientos es un poco mejor que el otro), un resultado igual a 0.5 implica un tamaño de efecto medio (uno de los tratamientos es claramente mejor que el otro) y un resultado igual a 0.8 implica un tamaño de efecto alto (uno de los tratamientos es mucho mejor que el otro) [4]. El tamaño de efecto estimado mediante WMD es simétrico respecto de los tratamientos. Un valor positivo indica que Y^E es preferible a Y^C , mientras que un valor negativo significa lo contrario.

Para mayores detalles de cómo aplicar las formulas indicadas remitirse a [4].

2.1.2 Response Ratio Paramétrico (PPR)

El PRR consiste en estimar un índice de efecto, o Ratio, entre dos tratamientos mediante el cociente de ambas medias [20]. Este cociente estima la proporción de mejora existente entre ambos tratamientos [21]. Así, por ejemplo, un ratio de 1.3 indicará que el

tratamiento principal es un 30% mejor que el secundario, o un ratio de 1 indicará que no hay diferencias en el desempeño de ambos tratamientos.

La aplicación del método es similar a WMD. Primeramente se debe estimar el Ratio de cada uno de los experimentos ($RR = Y^E / Y^C$) y luego, en base a estos, se estima el Ratio global mediante un promedio ponderado de los ratios individuales:

$$L^* = \frac{\sum_{i=1}^k W_i^* L_i}{\sum_{i=1}^k W_i^*} \quad \begin{array}{l} L^* \text{ es el efecto global} \\ L_i \text{ es el efecto de cada estudio} \\ W_i \text{ es el factor de peso} = 1/v \end{array} \quad (6)$$

Donde cada estudio es ponderado en base a la inversa de su varianza:

$$v = \frac{S^{2E}}{n^E Y^E} + \frac{S^{2C}}{n^C Y^C} \quad \begin{array}{l} v \text{ es el error típico} \\ S^{2's} \text{ son las varianzas de los estudios} \\ Y's \text{ son las medias de los estudios} \\ n's \text{ son las cantidades de sujetos} \end{array} \quad (7)$$

Para que la combinación de un conjunto de estudios sea más precisa se incorporó al método el logaritmo natural, el cual aplicado a los efectos de los estudios individuales permite linealizar los resultados y normalizar su distribución. Para mayores detalles de cómo aplicar las formulas indicadas remitirse a [21].

2.1.3 Conteo de Votos Estadístico (SVC)

El SVC es un método que requiere muy poca información para poder ser aplicado. Solo precisa conocer si existe o no diferencia entre las medias de los tratamientos (a lo cual llamaremos “voto”) y la cantidad de sujetos experimentales utilizados en cada estudio (utilizado como ponderador del “voto”) [4]. En base a estos datos se realiza un proceso de inferencia estadística con el objeto de determinar que tamaño de efecto (en general seleccionado de una lista que va desde -0,5 a 0,5) tiene la mayor probabilidad de ser el tamaño de efecto real que se hubiera estimado mediante WMD si se contara con todos los datos para poder hacerlo. La función principal de estimación es:

$$L(\delta | X_1, \dots, X_i) = \frac{\sum_{i=1}^k \left\{ X_i \ln \left[1 - \phi \left(-\sqrt{\tilde{n}} \delta \right) \right]_+ \right\}}{\sum_{i=1}^k \left\{ (1 - X_i) \ln \phi \left(-\sqrt{\tilde{n}} \delta \right) \right\}} \quad \begin{array}{l} L(\delta/X_1, \dots, X_n) \text{ es la probabilidad de tamaño de efecto} \\ \delta \text{ es el tamaño de efecto a testear} \\ X_i \text{ es el valor del voto de cada estudio} \\ \tilde{n} = (n^E + n^C) / (n^E * n^C) \text{ donde } n's \text{ son las cantidades de} \\ \text{sujetos experimentales de cada estudio} \end{array} \quad (8)$$

Para mayores detalles de cómo aplicar las formulas indicadas remitirse a [4].

2.1.4 Response Ratio No Paramétrico

Esta versión del RR es similar a la versión paramétrica, siendo su principal diferencia la forma en que pondera a los estudios. En lugar de utilizar la inversa de la varianza, el NPRR utiliza la cantidad de sujetos experimentales [21]:

$$v = \frac{n_C + n_E}{n_E n_C} + \frac{\text{Ln}(RR^2)}{2(n_C + n_E)} \quad \begin{array}{l} v \text{ es el error típico} \\ n's \text{ son las cantidades de sujetos} \\ RR \text{ es el Ratio} \end{array} \quad (9)$$

La principal ventaja de este método, desde el punto de vista de su aplicación, consiste en no requerir conocer las varianzas de los tratamientos ni requerir que exista normalidad y homeosticidad, lo cual aumenta enormemente sus probabilidades de aplicación. Para mayores detalles de cómo aplicar las formulas indicadas remitirse a [21].

2.2 Limitaciones de los métodos de Meta-Análisis

Los métodos presentados en la sección anterior (WMD, PRR, SVC, NPRR), aunque tienen el mismo objetivo (obtener una estimación sintética del tamaño de efecto entre dos tratamientos), no son en absoluto equivalentes. Se diferencian principalmente en dos aspectos: (1) el uso de parámetros poblacionales muestrales vs. estimaciones de los mismos y (2) la habilidad para incorporar mayor o menor número de estudios primarios.

Ambos aspectos están íntimamente relacionados. Debido a las carencias de reporte habituales en los experimentos de SE, solo unos pocos estudios publican medias, varianzas y tamaños muestrales que exigen métodos como WMD o PRR. Los métodos menos exigentes (NPRR, SVC), que no exigen la publicación de todos los parámetros, permiten incluir un mayor número de estudios, pero a costa de estimar los parámetros muestrales, perdiendo precisión durante este proceso.

Es importante destacar que para valores el análisis de estudios experimentales de *gran tamaño*, los métodos de meta análisis poseen distribuciones normales y parecidas precisiones. Sin embargo, en la práctica usual de la IS, los tamaños de los estudios experimentales no son *grandes*, sino *pequeños* y *bastante pequeños* y además en general la cantidad de estudios disponibles tampoco es *grande*. Por este motivo, la discrepancia entre las distribuciones teóricas y reales de WMD, PRR, SVC y NPRR (que no dejan de ser estimadores estadísticos) puede ser importante y un uso poco cuidadoso llevar a conclusiones erróneas. Si bien estos aspectos fueron tratados en los trabajos de Lajeunesse [14] y Friedrich [15], no fue hecho con el detalle necesario que requiere la IS en este momento.

3. Metodología

El objetivo de este trabajo es desarrollar un proceso de simulación, complementario a los anteriores [14, 15], para evaluar el desenvolvimiento de los cuatro métodos de Meta-Análisis (WMD, PRR, SVC y NPRR) de forma exhaustiva (variando: los tamaños de efectos y la cantidad de experimentos, la cantidad de sujetos por experimento), en un contexto experimental como el que hoy día presenta la SE. Donde, si bien la cantidad de experimentos ha venido creciendo significativamente en los últimos años (pasando de

aproximadamente tres estudios experimentales por año a principio de los 90' a más de veinte estudios por año, publicados en los principales congresos y revistas, a principio del años 2.000 [22]), todavía existen pocos experimentos y los mismos, en general, utilizan pocos sujetos experimentales (por ejemplo: [23] utiliza 4 sujetos experimentales) y tienen falencias en sus reportes (es común que no se publiquen las varianzas de los tratamientos analizados como sucede por ejemplo en: [24]).

De forma similar a como se hizo en [14] y [15], para desarrollar el proceso de simulación utilizaremos la técnica de Monte Carlo. La simulación de Monte Carlo es una técnica que combina conceptos estadísticos (muestreo aleatorio) con la capacidad que tienen los ordenadores para generar números pseudo-aleatorios siguiendo una distribución de probabilidad normal. En este contexto, se utilizó esta técnica para simular los valores que hubieran generado los distintos sujetos en la aplicación de los tratamientos, en base a los cuales, luego, se estimaron la media y la varianza de cada experimento.

El primer paso del desarrollo del proceso de simulación consiste en definir los valores poblacionales a partir de los cuales se obtienen los valores de la muestra para simular el Meta-Análisis. Para ello se siguieron las recomendaciones de los anteriores trabajos de simulación [14] y [15] e información relevante de cómo son típicamente los estudios hechos en el ámbito de la. Los tamaño de efecto poblacional a analizar son los mismo que se definen en [21] bajo (0,2), medio (0,5) y alto (0,8), mas la incorporación del tamaño de efectos muy alto (1,2), ya que se ha observado que varios estudios [25; 26] hechos en SE tienden a dar tamaños de efectos muy altos.

La media poblacional del tratamiento secundario (μ^c) es fijada en 100 y los desvíos estándar como se hizo en [21] son fijado en los siguientes porcentajes respecto de la media de dicho tratamiento: 10% al cual llamaremos varianza baja; 40% al cual llamaremos varianza media; y 70% al cual llamaremos varianza alta. Por su parte la media poblacional del tratamiento principal se estimará de la siguiente forma $\mu^E = 100 + \delta * \sigma$ y el ratio poblacional que se utilizará para validar los resultados que generen el RR paramétrico y no paramétrico será estimado: $RR = \mu^E / \mu^c$.

Por otra parte, la cantidad de experimentos a agregar en cada proceso de agregación irá desde 2 a 10 incrementándose de dos en dos, por considerar que el contexto experimental de la SE no aporta hoy día muchos experimentos potencialmente agregables mediante Meta-Análisis, tal y como puede comprobarse en las revisiones sistemáticas hechas hasta el momento [9; 10]. De igual modo, consideramos un número reducido de sujetos por experimento, en el rango de 4 – 20.

Las variables respuesta resultado de la simulación serán también las utilizadas en los estudios de Lajenouse [14] y Friedrich y colegas [15], esto es, el error de tipo I y II ó, para ser mas exacto, su inversa: (1- α) o precisión y (1 - β) o poder estadístico. Estas variables son adecuadas para nuestro propósito ya que determinan cuantas veces se equivoca un método de Meta-Análisis a la hora de determinar si existe (error de tipo I) o no (error de tipo II) una diferencia significativa entre dos tratamientos. Por último, siguiendo las

recomendaciones de [14] para cada combinación de valores de las variables se construirán 1.000 simulaciones, tras lo cual se calcularán los valores de las variables respuesta.

4. Resultados

Los resultados vinculados a la fiabilidad indican el porcentaje de veces que el intervalo de confianza estimado (a un nivel de $\alpha = 0.05$) contuvo el valor del tamaño de efecto poblacional, mientras que los resultados vinculados a la potencia estadística indican el porcentaje de veces que dicho intervalo de confianza no contuvo el valor 0 para los métodos WMD y SVC y el valor 1 para los métodos RR paramétricos y no paramétricos. Para facilitar la comprensión de las mismas, se han resaltado las celdas en las cuales los porcentajes estimados superaban al valor mínimo fijado, $1 - \alpha = 95\%$ para la fiabilidad y $1 - \beta = 80\%$ (el cual es el valor típicamente recomendado [27]) para la potencia estadística.

Respecto del desempeño de cada uno de los métodos podemos decir que:

- Es fiable utilizar el método WMD en contextos experimentales donde los tamaños de efecto poblacionales son bajos o medios, siendo su condición óptima de aplicación cuando los efectos son medios y el conjunto de experimentos a agregar superen a los 112 sujetos experimentales. Cuando los efectos poblacionales son altos o muy altos, el método tiende a perder fiabilidad sobre todo cuando se incrementa la cantidad de experimentos y la cantidad de sujetos experimentales.
- Es aconsejable utilizar el método PRR, siempre y cuando los estudios a agregar posean más de 4 sujetos experimentales. El método mostró ser robusto ante los cambios en la varianza, tamaños de efecto y cantidad de experimentos a agregar. Su condición óptima de aplicación varía en función del tamaño de efecto poblacional y la cantidad de sujetos experimentales que los estudios totalicen, observando que: para efectos muy altos se requieren por lo menos 80 sujetos experimentales, para efectos altos se requieren como mínimo 100 sujetos experimentales y para un efecto medio se requieren como mínimo 140 sujetos experimentales, para que el método posea fiabilidad y potencia estadística.
- Es fiable utilizar el método SVC, solo cuando el tamaño de efecto es medio se cuenta con experimentos que totalicen más de 80 sujetos experimentales. Su falta de fiabilidad es compensada en parte con su alta potencia, pero se debe tener mucho cuidado con el uso del mismo sobre todo en contextos experimentales donde el tamaño de efecto poblacional es bajo. En contextos de tamaños de efectos altos, la pérdida de fiabilidad es compensada en parte con la alta potencia estadística.
- El método NPRR ha sido el método más fiable de todos los analizados. Su mayor problema está dado por la baja potencia estadística que se acentúa en contextos donde la población tiene baja varianza. Esto se debe en parte a que en contexto de baja varianza no se requiere que la diferencia entre las medias sea excesiva para que el efecto sea alto. Su condición óptima de aplicación varía en función de la varianza poblacional, el tamaño de efecto poblacional y la cantidad de sujetos experimentales que los estudios totalicen observando que: para varianzas poblacionales medias y tamaños de efecto

poblacionales altos o muy altos se requieren como mínimo 100 sujetos experimentales, para varianzas poblacionales altas con tamaños de efecto poblacionales muy altos se requieren como mínimo 48 sujetos experimentales, para efectos poblacionales medios se requieren como mínimo 80 sujetos experimentales y para efectos poblacionales altos se requieren como mínimo 16 sujetos experimentales, para que el método posea fiabilidad y potencia estadística.

5 Conclusión

A modo de conclusión preliminar podemos decir que dentro de los parámetros normales que hoy presenta la Ingeniería de Software Empírica [5] el método WMD ha mostrado comportarse de forma confiable, por lo que no es necesario utilizar el método PRR como método alternativo al mismo. Por otra parte, en los casos en que los reportes experimentales no sean completos, el método NPRR mostró un comportamiento mucho más fiables que el SVC que, en general, no dio buenos resultados.

No obstante esto, si se trabaja en un entorno donde los tamaños de efecto son altos, el contexto cambia drásticamente, ya que aquí el método WMD deja de ser fiable, lo cual implica que los tamaños de efectos estimados pueden no ser correctos, por tal motivo el método PRR, que si ha mostrado ser fiable cuando los tamaños de efecto son altos, se convierte en el método mas recomendable cuando los reportes son completos, mientras que el método NPRR sigue siendo el mejor método cuando los reportes no son completos.

6 Referencias

1. Goodman C.; 1996; Literature Searching and Evidence Interpretation for Assessing Health Care Practices; SBU; Stockholm.
2. Basili, V. R., Green, S., Laitenberger, O., Lanubile, F., Shull, F., Sörumgård, S., Zelkowitz, M.; 1996; *The empirical investigation of perspective-based reading*, International Journal on Empirical Software Engineering, Vol. 1, No. 2; pp. 133–164.
3. Miller, J; 1999: Can Results from Software Engineering Experiments be Safely Combined? IEEE METRICS, 152-158
4. Hedges, L.; Olkin, I.; 1985; *Statistical methods for meta-analysis*. Academic Press.
5. Tonella P., Torchiano M., Du Bois B., Systä T.; 2007; *Empirical studies in reverse engineering: state of the art and future*
6. Sjöberg, D.; 2005; A survey of controlled Experiments in Software Engineering; ; IEEE Transactions on Software Engineering; Vol 31 Nro. 9
7. Biffi, S.; Halling, M.; Köszegi S.; 2003; *Investigating the Accuracy of Defect Estimation Models for Individuals and Teams Based on Inspection Data*; Proceedings of the 2003 International Symposium on Empirical Software Engineering (ISESE'03)

8. Fusaro, P., Lanubile, F., Visaggio, G.; 1997; *A replicated experiment to assess requirements inspection techniques*; Empirical Software Engineering, 2, 39-57.
9. Juristo N., Moreno A., Vegas S.; 2004; Towards building a solid empirical body of knowledge in testing techniques. ACM SIGSOFT Software Engineering Notes (SIGSOFT) 29(5):1-4
10. Jørgensen, M.; 2004; *A Review of Studies on Expert Estimation of Software Development Effort*. Journal of Systems and Software. (70): 1-2, pp. 37-60.
11. Dyba, T., Aricholm, E.; Sjöberg, D.; Hannay J.; Shull, F.; 2007; Are two heads better than one? On the effectiveness of pair programming. IEEE Software;12-15.
12. Shercliffe, R.; Stahl, W.; Tuttle, M.; 2009; *The Use of Meta-analysis in Psychology*; Theory & Psychology, Vol. 19, No. 3, 413-430 (2009)
13. Gurevitch, J. and Hedges, L.; 2001; *Meta-analysis: Combining results of independent experiments*. Design and Analysis of Ecological Experiments (eds S.M. Scheiner and J. Gurevitch), pp. 347–369. Oxford University Press, Oxford.
14. Lajeunesse, M & Forbes, M.; 2003; *Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques*. Ecology Letters, 6: 448-454.
15. Friedrich, J, Adhikari, N; Beyene, J; 2008; The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: A simulation study; BMC Medical Research Methodology
16. Cochrane; 2008; Curso Avanzado de Revisiones Sistemáticas; www.cochrane.es/?q=es/node/198
17. Glass, G; 1976; Primary, secondary, and meta-analysis of research. Educational Researcher 5: 3-8
18. Borenstein, M.; Hedges, L; Rothstein, H.; 2007; Meta-Analysis Fixed Effect vs. random effect; www.Meta-Analysis.com
19. Kitchenham, B. A.; 2004; *Procedures for performing systematic reviews*. Keele University; TR/SE-0401. Keele University Technical Report.
20. Worn, B.; Barbier, E.; Beaumont, N.; Duffy, J.; Folke, C; Halpern, B.; Jackson, J.; Lotze, H.; Micheli, F.; Palumbi, S.; Sala, E.; Selkoe, K.; Stachowics, J.; Watson, R; 2007; Supporting Online Material: Impacts of biodiversity loss on ocean ecosystem services.
21. Miguez, E. & Bollero, G; 2005; Review of Corn Yield Response under winter cover cropping systems using Meta-Analytic Methods; Crop Science Society of America
22. Sjöberg, D.; 2005; A survey of controlled Experiments in Software Engineering; ; IEEE Transactions on Software Engineering; Vol 31 Nro. 9
23. Burton, A., Shadbolt, N., Rugg, G. y Hedgecock, A.; 1990. *The Efficacy of Knowledge Elicitation Techniques: A Comparison Across Domains and Level of Expertise*. Knowledge Acquisition 2(2): 167-178.
24. Denger, C; Ciolkowski M; Lanubile, F; *Does Active Guidance Improve Software Inspections? A Preliminary Empirical Study*; 2004; Proceedings of the IASTED

International Conference SOFTWARE ENGINEERING February 17-19, 2004, Innsbruck, Austria; 408-413

25. Corbridge, C., Rugg, G., Major, P., Shadbolt, N. y Burton, A. 1994. *Laddering: Technical and Tool in Knowledge Acquisition*. Department of Psychology, University of Nottingham; Nottingham NG7 2RD.
26. Woody, J.; Will, R.; Blanton, J.; *Enhancing Knowledge Elicitation using the Cognitive Interview*; Expert system with application; 1996; Vol. 10 N. 1
27. Cohen, J.; *Statistical Power Analysis for the Behavioral Sciences*. (2nd ed.) 1988. ISBN 0-8058-0283-5.



 **Alfaomega Grupo Editor**