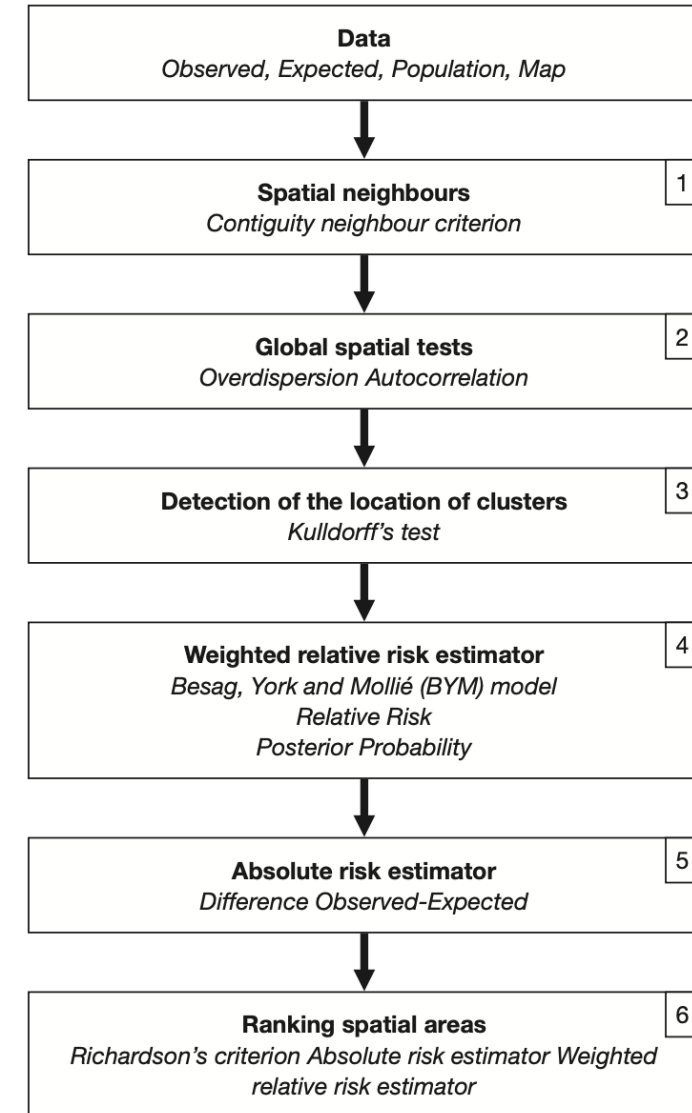


TALLER DE TRABAJO ONLINE

Detección y priorización de áreas de riesgo en enfermedades crónicas

Metodología I

1. Estrategia general
2. Diseño del pipeline de análisis
3. Ejemplos de aplicación
4. Pipeline
 - 4.0. Datos
 - 4.1. Vecindad espacial
 - 4.2. Test globales
 - 4.3. Detección de clusters
 - 4.4. Estimadores de riesgo relativo
 - 4.5. Estimador de riesgo absoluto
 - 4.6. Ranking



1. Estrategia general

- Análisis espaciales combinando estimadores relativos (SMR, RR, PP) y absolutos (casos observados y esperados).
- Objetivo: **identificar, ordenar y priorizar** áreas con mayor riesgo de una enfermedad.
- Zona de estudio dividida en **unidades espaciales no solapadas** (p. ej., municipios).
- Requisitos iniciales por área:
 - Casos observados
 - Casos esperados
 - Población

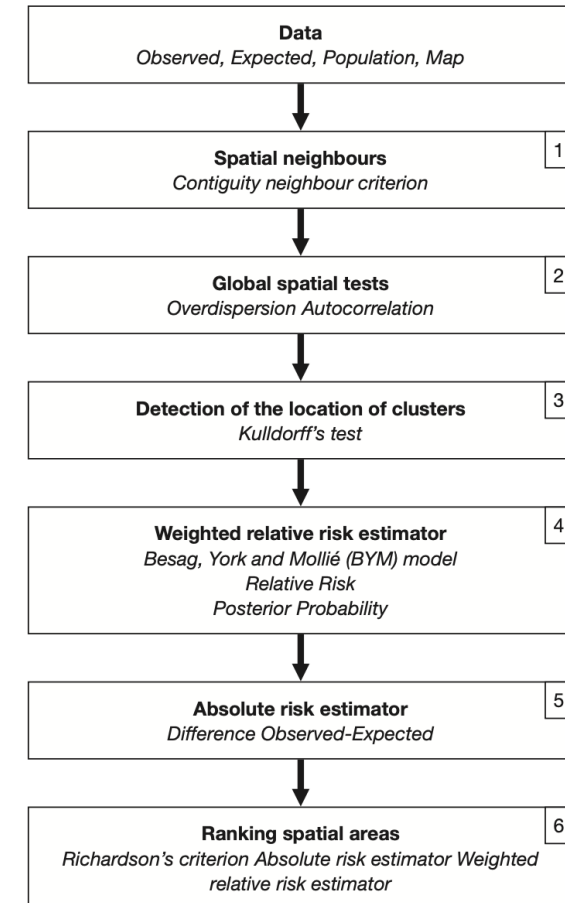
2. Diseño del pipeline de análisis

- **Pipeline integra:**

- Cálculo de riesgos relativos y absolutos.
- Evaluación de precisión de estimadores.
- Criterios y estrategias para rankear unidades espaciales según su riesgo epidemiológico.

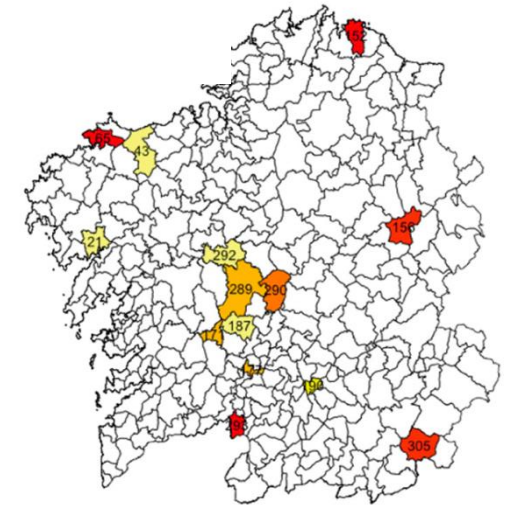
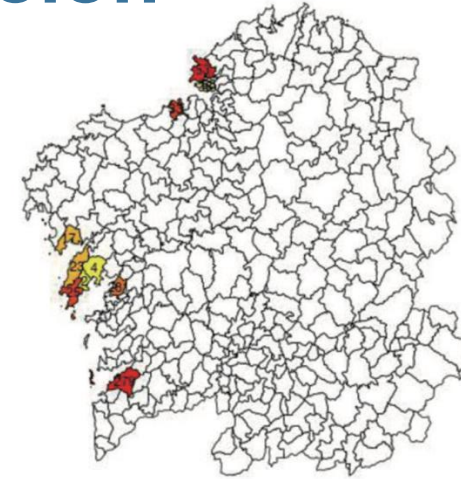
- **El pipeline incluye:**

- Evaluación preliminar (screening espacial).
- Tests estadísticos.
- Detección de clusters.
- Priorización final.

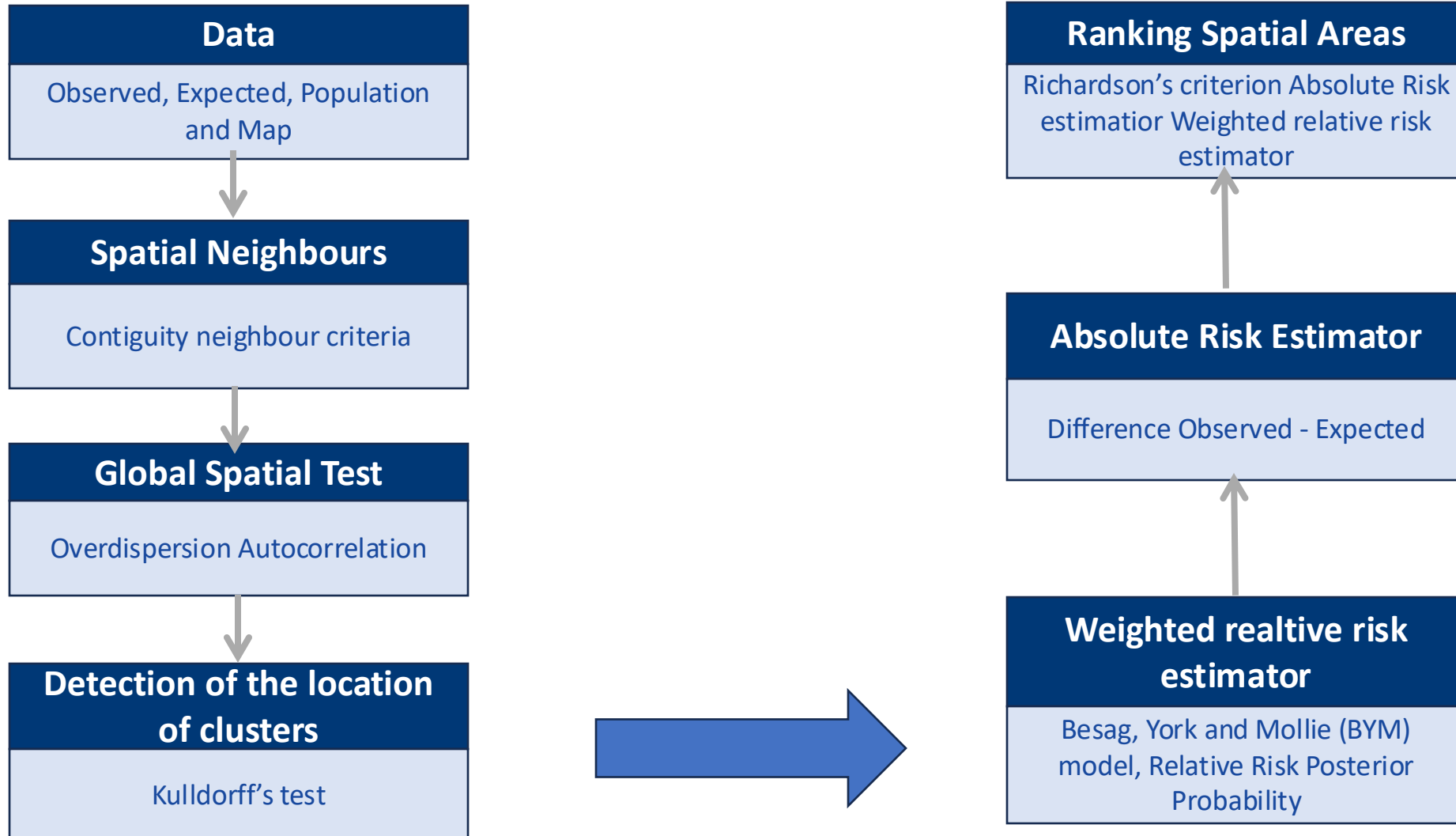


3. Ejemplos de aplicación

- **La metodología se evaluó en el artículo usando:**
 - Datos simulados, para ilustrar el funcionamiento interno.
 - Datos reales: mortalidad por cáncer de pulmón a nivel municipal.
- **Resultados:**
 - Identificación de zonas de interés epidemiológico.
 - Priorización de municipios para vigilancia.
 - Validación del pipeline y su utilidad en salud pública.



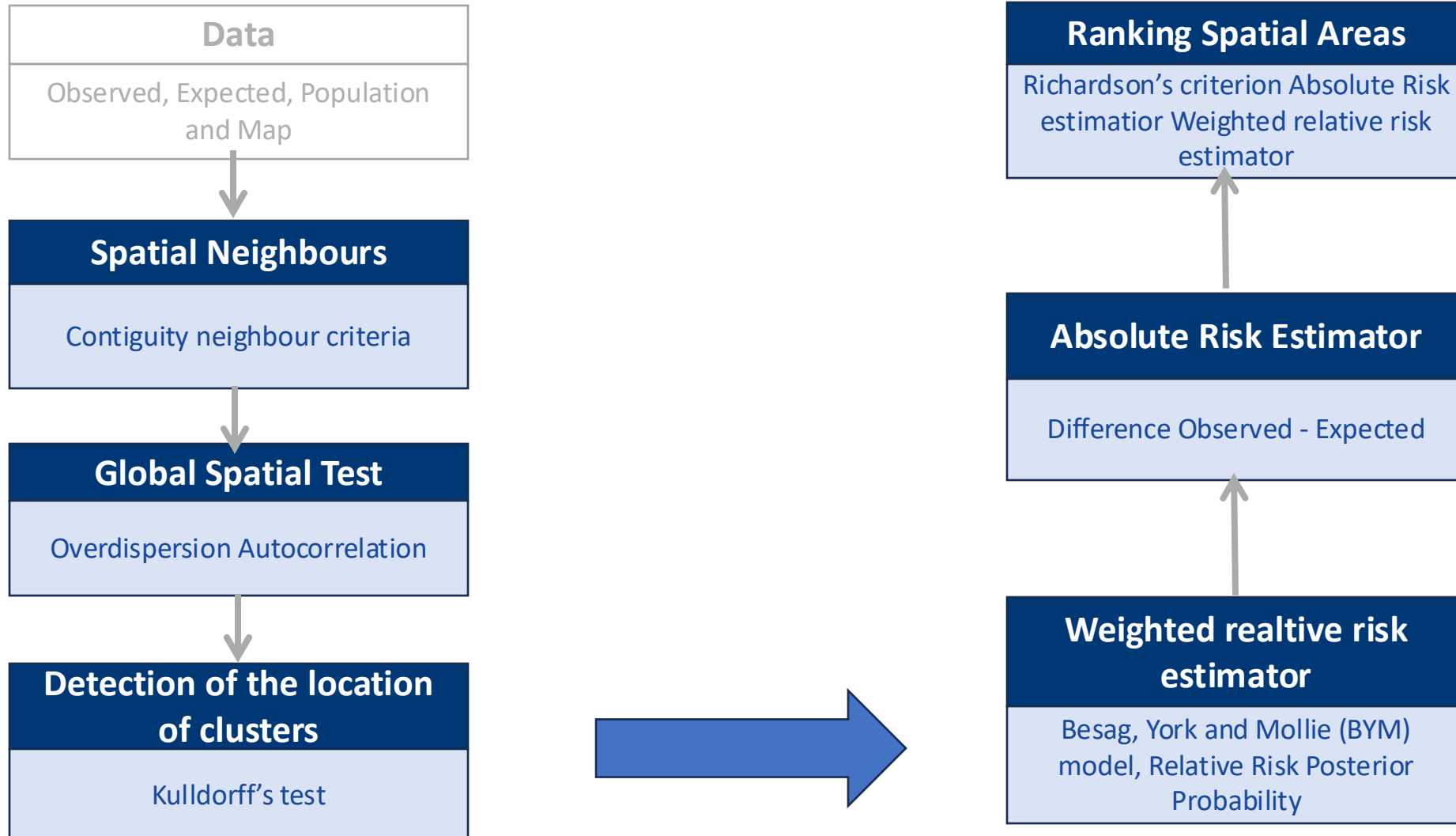
4. Pipeline



4.0. Datos

- El análisis se basa en: casos observados, **casos esperados** y población en unidades con fronteras definidas.
- En este tipo de datos es fundamental considerar la autocorrelación espacial.
- La autocorrelación implica que lo que ocurre en un área puede estar relacionado con lo que ocurre en las áreas vecinas.
- Ignorarla puede producir estimaciones sesgadas, errores en la detección de clusters y falsos positivos.

4. Pipeline



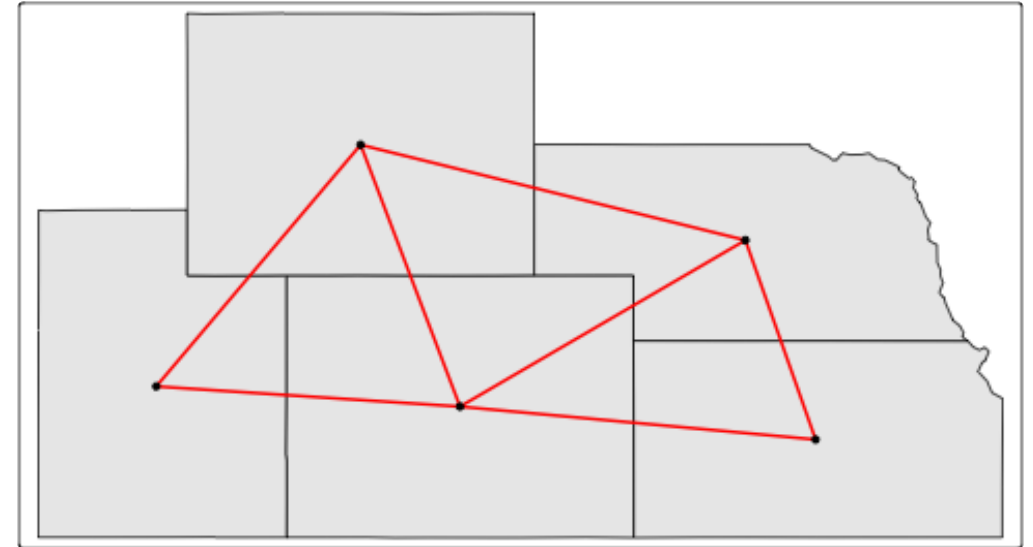
4.1. Vecindad espacial

- **Definir la vecindad espacial y construcción de la estructura de vecinos:**

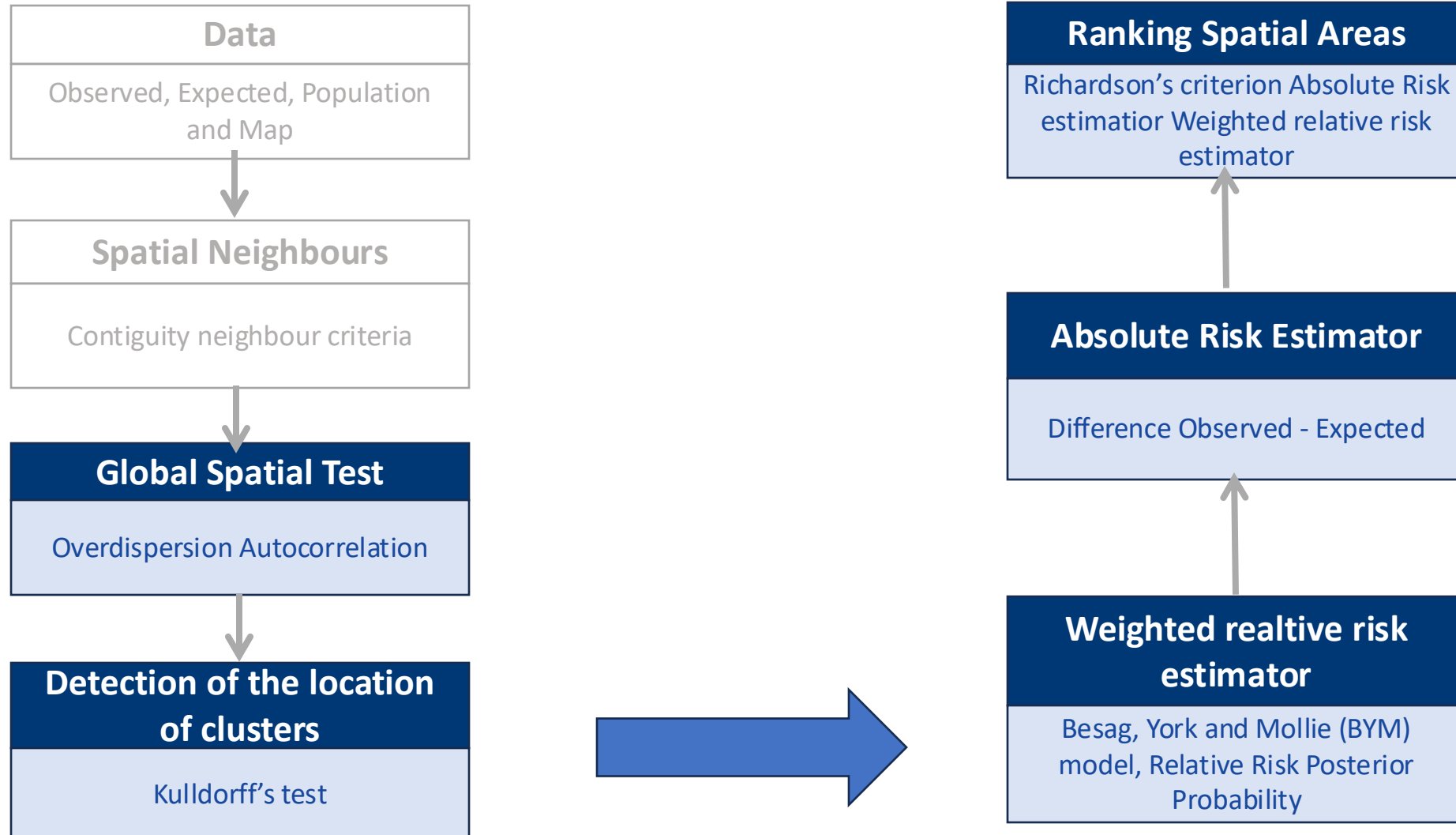
- Se utiliza un criterio estándar: contigüidad (polígonos que comparten borde o vértice son vecinos).
- Este criterio es apropiado para la mayoría de los análisis administrativos y epidemiológicos.
- En R, esto se implementa mediante la función `poly2nb()` del paquete `spdep`.
- Referencias:
 - Bivand RS, Pebesma E, Gómez-Rubio V. Applied Spatial Data Analysis with R. Springer; 2013.
 - Bivand R, Wong D. Comparing local indicators of spatial association using different neighbor definitions. Geographical Analysis. 2018.

4.1. Vecindad espacial

- **Por qué la estructura de vecindad es crucial:**
 - a) La estructura de vecindad afecta directamente a:
 - Tests globales (p. ej., Moran's I).
 - Tests locales y detección de clusters.
 - Modelos espaciales (CAR, BYM, etc.).
 - b) Una estructura inadecuada puede llevar a:
 - Subestimación o sobreestimación** de la autocorrelación.
 - Clusters artificiales creados por el método.
- **La contigüidad** es el estándar cuando se trabaja con áreas administrativas, pero existen alternativas según el fenómeno (distancia, k-vecinos, etc.).



4. Pipeline



4.2. Test globales

- **¿Por qué realizar análisis espaciales globales preliminares?**
 - Antes de aplicar modelos o construir rankings de áreas de riesgo, es esencial evaluar si los datos presentan patrones espaciales no aleatorios.
 - Esto permite evitar errores de interpretación y seleccionar modelos adecuados.
- **Tests estadísticos para evaluar:**
 - Sobredispersión: Diferencias globales entre casos observados y esperados que no encajan con un modelo Poisson.
 - Autocorrelación espacial: Determina si las áreas vecinas presentan riesgos similares.

4.2. Test globales

- **Utilidad de estos tests:**

- Escoger el modelo adecuado

- Sobredispersión \Rightarrow modelos NB o bayesianos con heterogeneidad
- Autocorrelación \Rightarrow modelos CAR/BYM

- Evitar conclusiones erróneas

- No confundir ruido con patrones espaciales reales
- Evitar rankings engañosos por azar

- Optimizar el procedimiento

- Identifica si existe señal suficiente para un cribado espacial
- Prioriza áreas antes de análisis más complejos

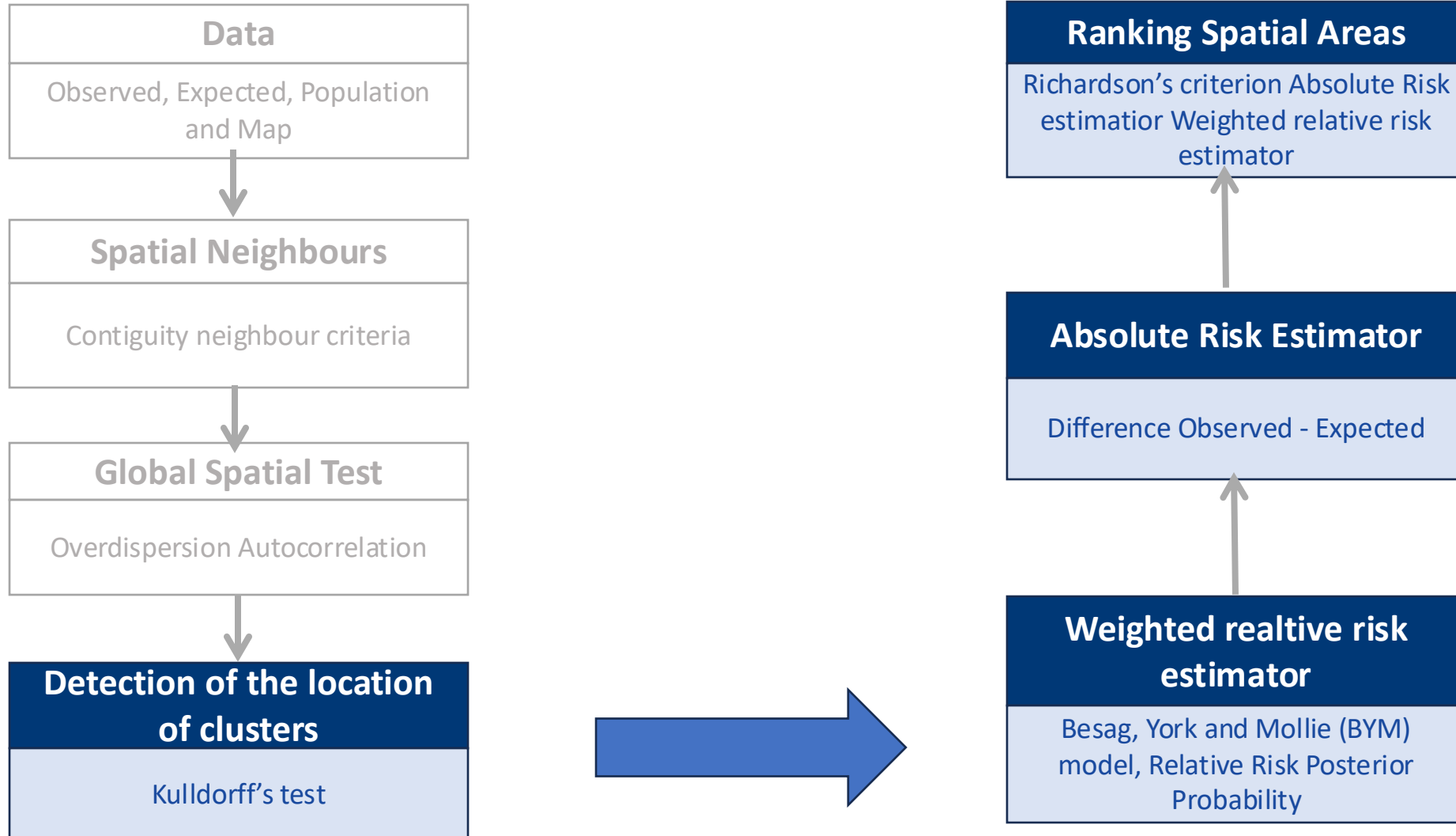
4.2. Test globales

- **Sobredispersión:**
 - Diferencias globales entre casos observados y esperados que no encajan con un modelo Poisson.
 - ¿Por qué es importante?
 - La sobredispersión indica heterogeneidad del riesgo entre áreas.
 - Puede deberse a factores estructurales (sistemas sanitarios distintos, focos ambientales...).
 - Justifica el uso de modelos bayesianos con término de heterogeneidad o modelos NB. (Bivand et al., 2013; Lawson, 2018)
 - Tests usados:
 - Chi-square test
 - Potthoff-Whittinghill test
 - Implementados en paquete DCluster: `achisq.test()` y `pottwhitt.test()`

4.2. Test globales

- **Autocorrelación espacial**
 - Determina si las áreas vecinas presentan riesgos similares.
 - **¿Por qué importa?**
 - La autocorrelación indica que el riesgo depende del espacio.
 - Justifica el uso de **modelos espaciales correlados (CAR/BYM)** frente a modelos independientes.
 - **Test aplicados**
 - **Moran's I** (global clustering)
 - **Tango's Maximised Excess Events test** (clustering por proximidad)
 - Implementados en el paquete DCluster: `moranI.test()`, `tango.test()`

4. Pipeline



4.3. Detección de clusters

- Complementa los tests de autocorrelación global (Moran/Tango).
- Se aplica el test de Kulldorff y Nagarwalla (1995) para detectar clusters espaciales.
- El test busca **clusters de exceso de casos** respecto a lo esperado, no clusters de riesgos similares.
- Basado en una ventana circular que se desplaza por el mapa.
- Motivo de uso:
 - Método bien establecido en epidemiología.
 - Fácil de implementar y computacionalmente eficiente.
- Limitaciones:
 - Asume clusters circulares → puede no detectar formas irregulares.
 - Sensible al parámetro de fracción máxima de población incluida en el cluster.

4.3. Detección de clusters

- **Implementación en R**

- Se usa la función `opgam()` del paquete `DCluster`.
- Parámetros típicos:
 - Modelo: Negative Binomial
 - Bootstrap: 99 réplicas
 - Significancia: 5%
 - Fracción de población: 15% para la creación de círculos
 - Solo se consideran clusters con $p < 0.05$.
 - La fracción de población puede ajustarse para explorar si la ubicación de los clusters cambia.
- Bivand et al., 2013

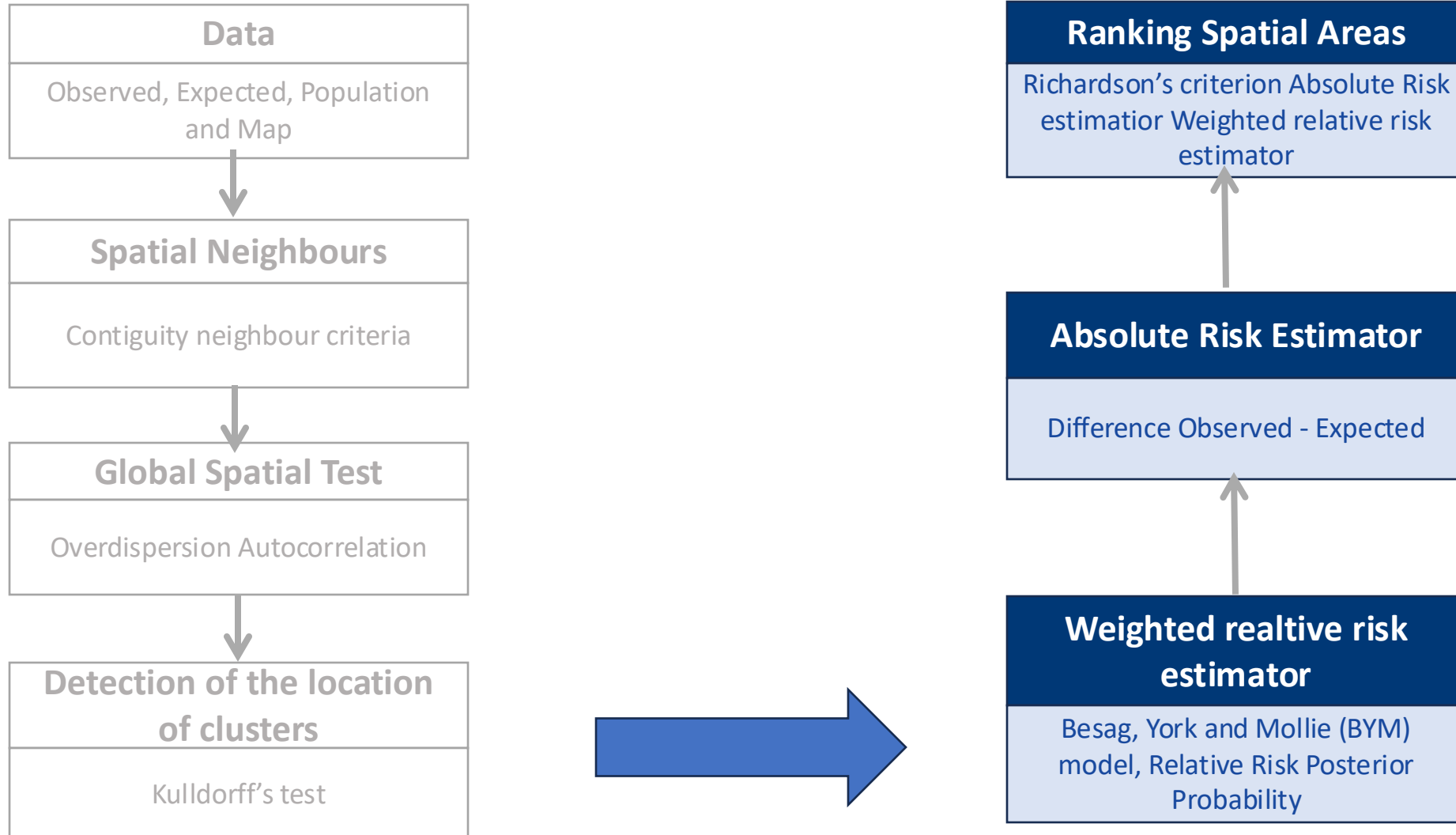
4.3. Detección de clusters

- Una vez detectados los clusters ($p < 0.05$), se **ordenan según el valor del estadístico** obtenido en el test de Kulldorff
- Este estadístico refleja cuán mayor es la evidencia de exceso de casos en comparación con lo esperado
 - Statistic: diferencia en log-verosimilitud entre el cluster detectado y el modelo nulo:

$$Statistic = \llbracket "LL" \rrbracket_{\text{"cluster"}} - \llbracket "LL" \rrbracket_{\text{"nulo"}}$$

- Se pueden obtener del análisis:
 - mapas donde se identifican los cluster (con su centro).
 - tabla resumen con los cluster significativos, Statistic , número de cluster...

4. Pipeline



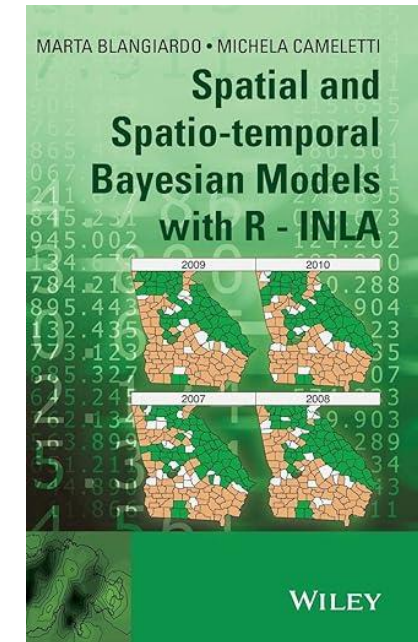
4.4. Estimadores de riesgo relativo

- Se calculan estimadores locales de riesgo relativo (SMR suavizadas o RR) por área usando modelos Bayesianos.
- Se obtiene también los intervalos de credibilidad de estas estimaciones de Riesgo y las probabilidades a posterior (PP) de que $RR > 1$
- Modelo BYM (Besag, York, Mollié):
 - Modelo de Poisson con:
 - Casos observados** como variable dependiente
 - Log-casos esperados** como offset
 - Dos términos de efectos aleatorios:
 - Efecto estructurado espacial** (autocorrelación por contigüidad)
 - Efecto heterogeneidad no espacial**
 - Ventaja: **permite suavizar SMR (RR)**, respetando correlación espacial y heterogeneidad local.
 - Referencia: Besag et al., 1991

4.4. Estimadores de riesgo relativo

- **Herramienta de inferencia bayesiana:**

- **INLA (Integrated Nested Laplace Approximation):**
- Calcula estimaciones bayesianas de manera **rápida y confiable**
- Menor costo computacional que MCMC tradicional
- Implementación en R: **inla()** del paquete **R-INLA**
 - Estimación de parámetros con Laplace simplificada
 - Priors por defecto adecuados para RR y heterogeneidad
- **Referencias:** Rue et al., 2009; Blangiardo & Cameletti, 2015



J. R. Statist. Soc. B (2009)
71, Part 2, pp. 319–392

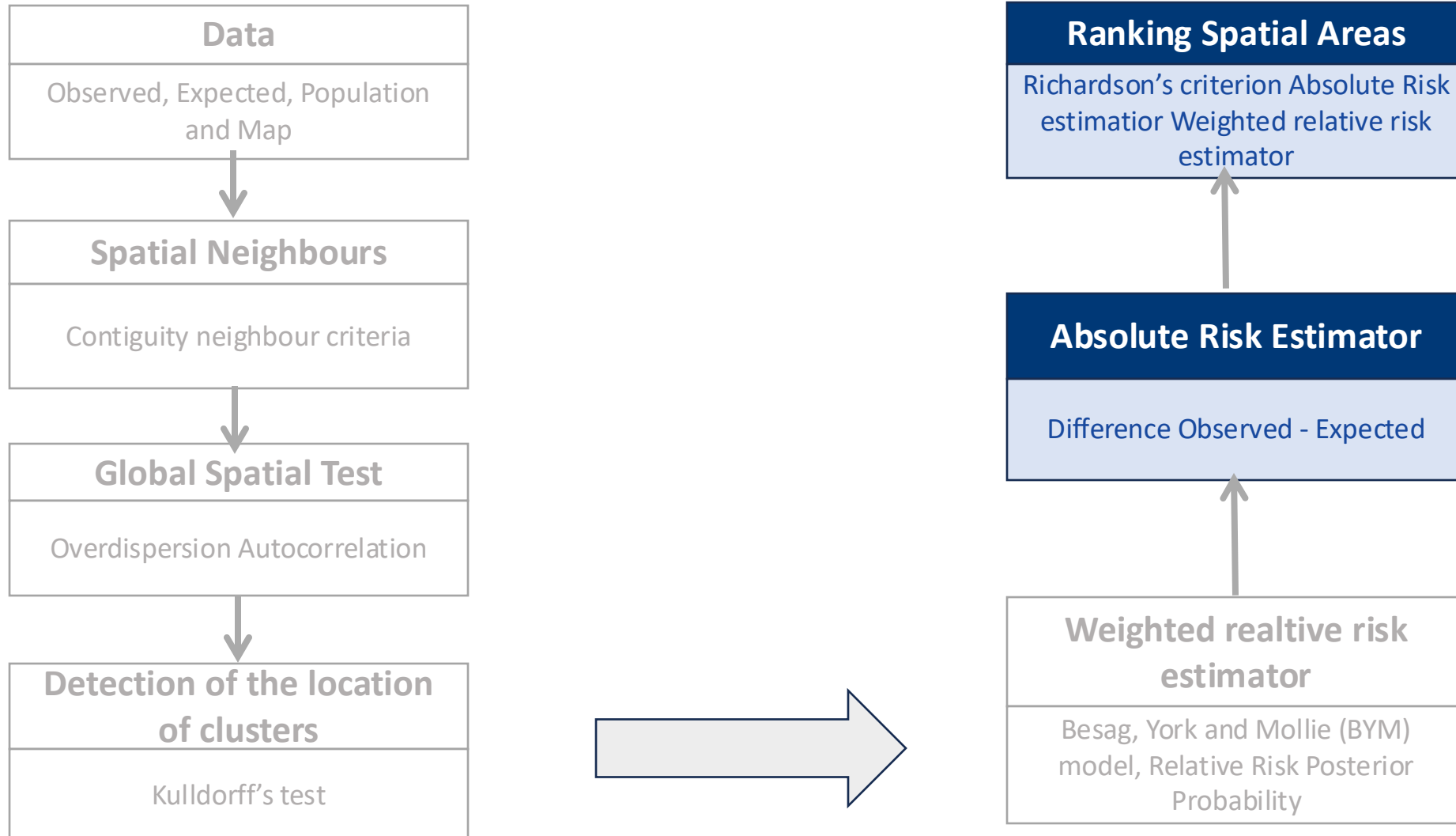
Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations

Håvard Rue and Sara Martino

4.4. Estimadores de riesgo relativo

- **RR ponderado (Weighted RR, WRR)**
 - Penalización de RR por precisión
 - Se calcula **WRR (score en el pipeline del ejemplo práctico) = $RR \times PP$**
 - Beneficio:
 - Penaliza áreas con **alta incertidumbre** (PP baja)
 - Mejora discriminación de áreas con riesgo significativo
 - Evita que áreas con pocos casos y alta variabilidad influyan en el ranking final
- **Resultado final:**
 - Una **medida de riesgo robusta**, combinando magnitud y confiabilidad de la estimación.
 - Lista de áreas **ordenadas según WRR**, útil para priorizar vigilancia e intervenciones.

4. Pipeline



4.5. Estimador de riesgo absoluto

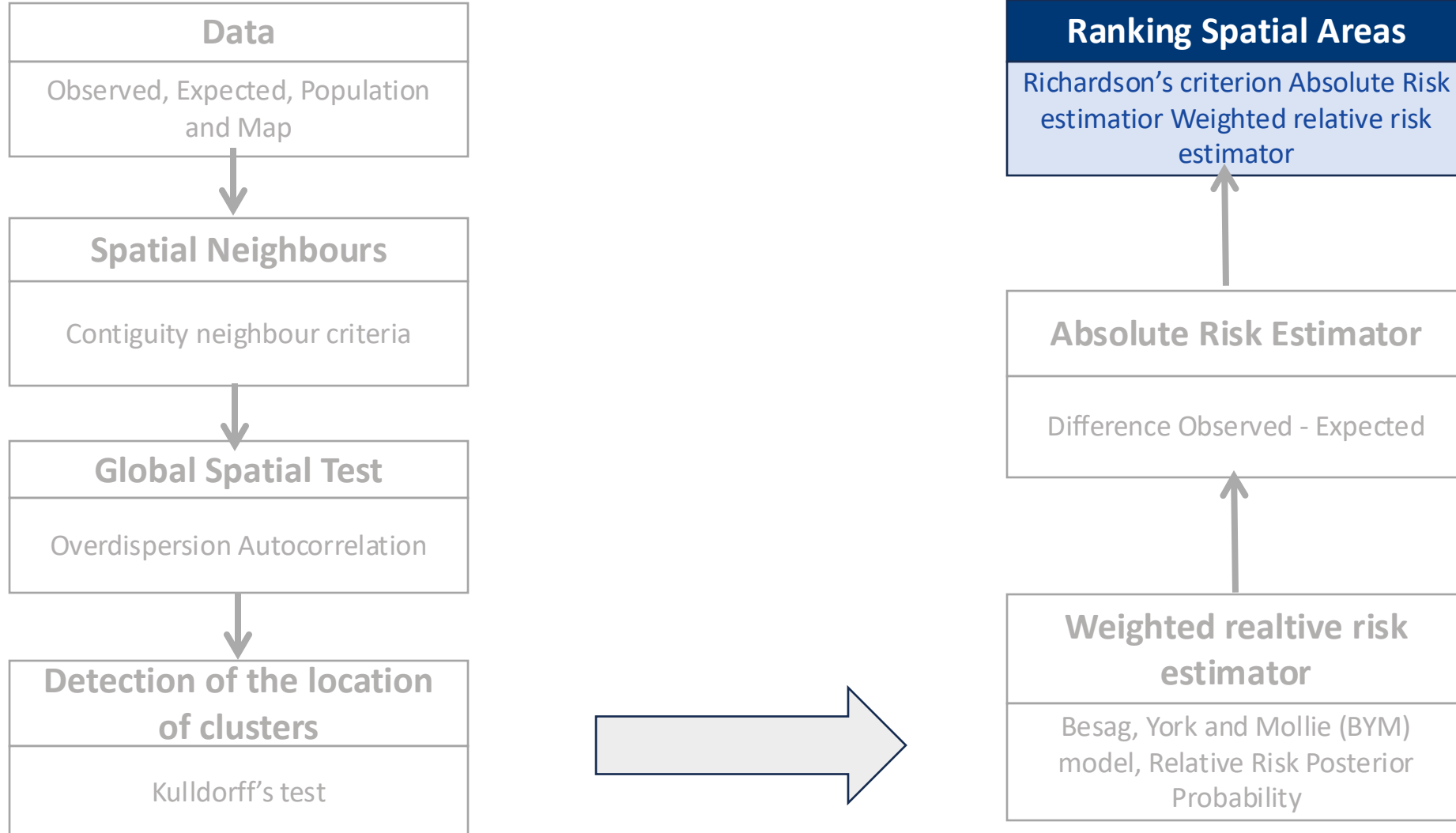
- **Diferencia observada-Esperada (DOE)**
 - Se Calcula el **exceso absoluto de casos** en cada área:
 - $DOE(\text{Diff_obs_exp}) = \text{Casos observados} - \text{Casos esperados}$
 - Complementa los **RR/SMR suavizados** que son relativos.
 - Expresado como **número de casos**, útil para la **gestión sanitaria**.
- **RR / SMR:** Riesgo relativo respecto al esperado. Comparación entre áreas con poblaciones diferentes
- **DOE:** Exceso absoluto de casos. Planificación de recursos, intervenciones y priorización de áreas

4.5. Estimador de riesgo absoluto

¿Por qué calcular DOE?

- Permite identificar **áreas con alto impacto absoluto**, independientemente de la población.
- Complementa los **RR relativos**, que pueden exagerar áreas pequeñas con pocos casos.
- Facilita la **priorización en salud pública**, asignación de recursos y planificación de intervenciones.
- Forma parte del **ranking final de áreas de riesgo**, combinando:
 - Magnitud relativa (RR/SMR suavizado)
 - Precisión (PP → WRR)
 - Magnitud absoluta (DOE)

4. Pipeline



4.6. Ranking

- **Paso previo al Ranking: Identificación de áreas de alto riesgo**
 - Selección según probabilidad posterior (PP)
 - Se seleccionan áreas con **PP > 80%** según **Richardson's criterion**.
 - Solo estas áreas se consideran para el ranking prioritario.
 - Las áreas con **PP ≤ 80%** reciben un **valor de ranking inferior**, indicando menor certeza de riesgo elevado.
- **Justificación:**
 - Garantiza que solo se prioricen áreas con **evidencia suficiente de exceso de riesgo**.
 - Evita sobreinterpretar áreas con pocos casos o alta incertidumbre.

4.6. Ranking

- **Ranking de áreas: Combinación de indicadores para priorizar áreas**
- **Ordenamiento de las áreas seleccionadas:**
 - **Primero por DOE:** magnitud absoluta de exceso de casos → áreas con mayor impacto en salud pública.
 - **Luego por WRR:** RR suavizado ponderado por probabilidad posterior → integra riesgo relativo y precisión.
 - Cada área recibe un **número de ranking**, empezando por **1 = más relevante**.
 - Áreas con $PP \leq 80\%$ → asignadas **después de las áreas seleccionadas**, mismo valor de ranking más alto + 1.