

## Nota para publicación

En el marco de la suscripción del acuerdo de entrega de información por parte de Patrick Ball al Departamento Administrativo Nacional de Estadística (DANE), en el que se recibieron las bases de datos anonimizadas resultado del proyecto *integración de datos y estimaciones de estadísticas en el conflicto armado*, el equipo técnico del DANE realizó ejercicios de revisión del proceso de anonimización implementado con el fin de determinar si existían riesgos adicionales de reidentificación no contemplados previamente. Estas bases de datos consistieron en cien réplicas de los datos generados por la metodología de imputación establecida para los componentes de homicidio, desaparición, secuestro y reclutamiento.

Para ello, se siguió una metodología de anonimización que consistió en determinar las variables sensibles, establecer llaves de reidentificación y determinar los registros únicos en función de estas llaves. Se realizaron dos escenarios: el Escenario 1 con la llave departamento y fecha del evento, y el Escenario 2, que incluyó además la variable municipio.

Tras identificar los registros únicos y analizar los riesgos en los registros sin intervención o no imputados, se determinó que había 1.952 personas en riesgo en el Escenario 1 y 35.003 personas en riesgo en el Escenario 2, excluyendo los registros fuera del periodo de difusión establecido por la Comisión para el Esclarecimiento de la Verdad (CEV).

En una segunda fase, se realizaron ejercicios de rompimiento utilizando bases de datos externas con una estructura similar a las de la CEV, como el Registro Estadístico Base de Población (REBP), el Registro Único de Víctimas (RUV) y las Estadísticas Vitales. Estos ejercicios confirmaron cruces efectivos entre los registros en los diferentes escenarios, permitiendo validar la precisión de los registros no imputados.

A partir de estos resultados, se presentaron recomendaciones al equipo de la Comisión de la Verdad para realizar intervenciones adicionales en las bases de datos anonimizadas. Estas intervenciones incluyeron la agrupación de la edad en grupos quinquenales, la exclusión de registros anteriores al periodo definido por la CEV y la perturbación del municipio y la fecha del evento en los registros de mayor riesgo. Estas intervenciones fueron implementadas por el equipo de Patrick Ball utilizando técnicas de imputación múltiple y validadas por el DANE, logrando así reducir significativamente los niveles de riesgo de reidentificación en las bases de datos anonimizadas que serán disponibles en datos abiertos.