# Supplements

In Section S1, we present preliminary definitions for this study. In Section S2, we give detailed proofs for the theoretical guarantees of our methods. In Section S3, we describe the generation procedures of the simulation data used in our experiments.

## S1. Preliminaries

### A. Linkage Analysis

$$
\begin{aligned}
\chi^2_{td} &:= \chi^2_{td}(h,i,j) = \frac{2(i-j)^2}{h}, \\
\chi^2_{hs} &:= \chi^2_{hs}(h,i,j) = \frac{(2i+2j-h)^2}{h}, \\
\chi^2_{total} &:= \chi^2_{total}(h,i,j) \\
&= \frac{(i-h/4)^2}{h/4} + \frac{(h-i-j-h/2)^2}{h/2} + \frac{(j-h/4)^2}{h/4} \\
&= \frac{2(i-j)^2}{h} + \frac{(2i+2j-h)^2}{h} = \chi^2_{td} + \chi^2_{hs}.
\end{aligned}
$$

### B. Differential Privacy

**Definition S1.** *($\epsilon$-Differential Privacy [1])*
*A randomized mechanism M is $\epsilon$-differentially private if, for all datasets D and $D'$ which differ in only one family and any $S \subset$range(M),*

$$\Pr[M(D) \in S] \leq e^{\epsilon} \cdot \Pr[M(D') \in S].$$

*1) Laplace Mechanism :*

**Definition S2.** *(Sensitivity for the Laplace Mechanism [2])*
*Let $\mathcal{D}^M$ be the collection of all datasets with $M$ SNPs, the sensitivity of a function $f : \mathcal{D}^M \to \mathbb{R}^d$ is*

$$\Delta f = \max_{D,D'} \|f(D) - f(D')\|_1,$$

*where $D, D' \in \mathcal{D}^M$ differ in a single family.*

For a statistic $f(D)$ obtained from the original dataset $D$, releasing $f(D) + b$ satisfies $\epsilon$-differential privacy when $b$ is random noise derived from a Laplace distribution with mean 0 and scale $\frac{\Delta f}{\epsilon}$.

*2) Exponential Mechanism:*

**Definition S3.** *(Sensitivity for the Exponential Mechanism [3])*
*Let $\mathcal{D}^M$ be the collection of all datasets with $M$ SNPs, the sensitivity of a score function $u : \mathcal{D}^M \times \{1, 2, \ldots, M\} \to \mathbb{R}$ is*

$$\Delta u = \max_{r} \max_{D,D'} |u(D,r) - u(D',r)|,$$

*where $r \in \{1, 2, \ldots, M\}$ and $D, D' \in \mathcal{D}^M$ differ in a single family.*

Following the above definition, we choose the mechanism $\mathcal{M}_u^{\epsilon}$ which has distribution

$$\mathcal{M}_u^{\epsilon} = \frac{\exp\left(\frac{\epsilon u(D,r)}{2\Delta u}\right)}{\sum_{s \in \{1,\ldots,M\}} \exp\left(\frac{\epsilon u(D,s)}{2\Delta u}\right)}.$$

Then, releasing $\mathcal{M}_u^{\epsilon}$ satisfies the definition of $\epsilon$-differential privacy.

In this study, we use the Shortest Hamming Distance (SHD) score as the score function, and the definition is shown below.

**Definition S4.** *(The SHD score [4])*
*Given a predefined threshold $c^* > 0$, the SHD score for $i$-th data $D_i \, (i = 1, 2, \ldots, M)$ is*

$$d_{\mathrm{SH}}(D_i, i) = \begin{cases} 0, \ (T_i \geq c^* \wedge \exists D_i', T_i' < c^*) \\ 1 + \min d_{\mathrm{SH}}(D_i', i), \ (T_i \geq c^* \wedge \nexists D_i', T_i' < c^*) \\ -1 + \max d_{\mathrm{SH}}(D_i', i), \ (T_i < c^*) \end{cases}$$

*where $T_i$ and $T_i'$ are the test statistics obtained from $D_i$ and $D_i'$, respectively, and $D_i, D_i' \in \mathcal{D}^M$ differ in a single family. For $i \in \{1, \ldots, M\}$, $d_{\mathrm{SH}}(D_i, i) = -\infty$.*

## S2. Proofs

### A. Laplace Mechanism

**Theorem S1.** *The sensitivity of the statistic $\chi^2_{td}$ for $n$ families with each family having two affected children is $\frac{16(n-1)}{n}$.*

*Proof.* The statistic $\chi^2_{td}$ can be expressed as a function

$$\chi^2_{td} : \mathcal{D} \longrightarrow \mathbb{R}_{\geq 0},$$

where $\mathcal{D} = \{(h,i,j) \in \mathbb{N} \mid 0 \leq i,j \leq h, i+j \leq h, 10 < h \leq 2n\}$.

The possible combinations of $(h,i,j)$ in one family are

$$
\begin{aligned}
(h,i,j) \;=\; & (0,0,0), (1,0,0), (1,0,1), (1,1,0), (2,0,0), \\
& (2,0,1), (2,0,2), (2,1,0), (2,1,1), (2,0,2).
\end{aligned}
$$

In the following, we divide the cases according to the amount of change in the value of $h$ between two neighboring datasets.

(I) When the value of $h$ does not change.

The changes in the values of $(i,j)$ are as follows:

$$(i,j) \leftrightarrow \begin{cases} (i-a, j+b) & (a \in \{0,1,2\}, b \in \{0,1,2\}) \\ (i \pm 1, j \pm 1) & (\text{double-sign corresponds}) \\ (i+a, j-b) & (a \in \{0,1,2\}, b \in \{0,1,2\}) \end{cases}.$$

(i) $(i,j) \leftrightarrow (i-a, j+b)$

$$\begin{aligned}&\chi^2_{td}(h,i,j) - \chi^2_{td}(h,i-a,j+b)\\ =\;& \frac{2(a+b)(2i-2j-a-b)}{h}\end{aligned} \tag{1}$$

If $a \geq b$, i.e., $(a,b) = (0,0)$, $(1,0)$, $(1,1)$, $(2,0)$, $(2,1)$, $(2,2)$, then $a \leq i \leq h$, $0 \leq j \leq h-a$, $i+j \leq h$. In this case,

$$\begin{aligned}& -h+2a \leq i-j \leq h\\ \Leftrightarrow\;& -2h+3a-b \leq 2i-2j-a-b \leq 2h-a-b.\end{aligned}$$

Therefore,

$$\begin{aligned}|(1)| &\leq \frac{2(a+b)(2h-a-b)}{h}\\ &\leq \frac{16(h-2)}{h} \leq \frac{16(2n-2)}{2n} = \frac{16(n-1)}{n}.\end{aligned}$$

If $a < b$, as in the case of $a \geq b$, $|(1)| \leq \frac{3(4n-3)}{n}$.
(ii) $(i,j) \leftrightarrow (i \pm 1, j \pm 1)$
The statistic is not changed since

$$\chi^2_{td}(h,i,j) = \chi^2_{td}(h, i\pm 1, j \pm 1).$$

(iii) $(i,j) \leftrightarrow (i+a, j-b)$
Same as the case (i).

(II) When $h$ changes to $h+1$.

The changes in the values of $(i,j)$ are as follows:

$$(i,j) \leftrightarrow \begin{cases}(i-a, j+b) & (a \in \{0,1\}, b \in \{0,1,2\})\\ (i+1, j+1) & \\ (i+a, j-b) & (a \in \{0,1,2\}, b \in \{0,1\})\end{cases}.$$

(i) $(i,j) \leftrightarrow (i-a, j+b)$

$$\begin{aligned}&\chi^2_{td}(h,i,j) - \chi^2_{td}(h+1,i-a,j+b)\\ =\;& \frac{2\{(i-j)^2 + 2h(a+b)(i-j) - h(a+b)^2\}}{h(h+1)}\end{aligned} \tag{2}$$

If $a \geq b$, i.e., $(a,b) = (0,0), (1,0), (1,1)$, then $a \leq i \leq h$, $0 \leq j \leq h-a$, $i+j \leq h$. In this case, $-h+2a \leq i-j \leq h$. Therefore,

$$\begin{aligned}|(2)| &\leq \frac{2\{-(a+b)^2 + 2h(a+b) + h\}}{h+1}\\ &\leq \frac{2(-4+5h)}{h+1} \leq \frac{2(10n-9)}{2n} = \frac{10n-9}{n}.\end{aligned}$$

If $a < b$, as in the case of $a \geq b$, $|(2)| \leq \frac{2(7n-8)}{n}$.
(ii) $(i,j) \leftrightarrow (i+1, j+1)$

$$\chi^2_{td}(h,i,j) - \chi^2_{td}(h+1,i+1,j+1) = \frac{2(i-j)^2}{h(h+1)} \tag{3}$$

Since $0 \leq i \leq h-2$, $0 \leq j \leq h-2$, $i+j \leq h$,

$$|(3)| \leq \frac{2(h-2)^2}{h(h+1)} \leq \frac{2(2n-3)^2}{(2n-1)2n} = \frac{(2n-3)^2}{n(2n-1)}.$$

(iii) $(i,j) \leftrightarrow (i+a, j-b)$
Same as the case (i).

(III) When $h$ changes to $h+2$.

The changes in the values of $(i,j)$ are as follows:

$$(i,j) \leftrightarrow \begin{cases}(i+a, j) & (a \in \{0,1,2\})\\ (i+1, j+1) & \\ (i, j+b) & (b \in \{0,1,2\})\end{cases}.$$

(i) $(i,j) \leftrightarrow (i+a, j)$

$$\begin{aligned}&\chi^2_{td}(h,i,j) - \chi^2_{td}(h+2,i+a,j)\\ =\;& \frac{2\{2(i-j)^2 - 2ah(i-j) - ha^2\}}{h(h+2)}\end{aligned} \tag{4}$$

Since $0 \leq i \leq h-a$, $0 \leq j \leq h-a$, $i+j \leq h$,

$$\begin{aligned}|(4)| &< \frac{2\{2(1+a)h - a^2\}}{h+2}\\ &< \frac{4(3h-2)}{h+2} \leq \frac{4(6n-8)}{2n} = \frac{4(3n-4)}{n}\end{aligned}$$

(ii) $(i,j) \leftrightarrow (i+1, j+1)$
Similarly to the case (II) (ii),

$$\left|\chi^2_{td}(h,i,j) - \chi^2_{td}(h+2,i+1,j+1)\right| \leq \frac{4(n-4)^2}{n(n-2)}$$

(iii) $(i,j) \leftrightarrow (i, j+b)$
Same as the case (i).

(IV) When $h$ changes to $h-1$ or $h-2$.

Same as the case (II) and (III).
From the above, the sensitivity of $\chi^2_{td}$ is $\frac{16(n-1)}{n}$. □

**Theorem S2.** *The sensitivity of the statistic $\chi^2_{hs}$ for $n$ families with each family having two affected children is $\frac{8(n-1)}{n}$.*

*Proof.* The statistic $\chi^2_{hs}$ can be expressed as a function

$$\chi^2_{hs} : \mathcal{D} \longrightarrow \mathbb{R}_{\geq 0},$$

where $\mathcal{D} = \{(h,i,j) \in \mathbb{N} \mid 0 \leq i,j \leq h, i+j \leq h, 10 < h \leq 2n\}$.

As in the Theorem S1, we divide the cases according to the amount of change in the value of $h$. In the following proof, we note that $0 \leq i+j \leq h$.

(I) When the value of $h$ does not change.

The change in $\chi^2_{hs}(h,i,j)$ is as follows:
(i) $(i,j) \leftrightarrow (i-2,j), (i-1,j-1), (i,j-2)$

$$\begin{aligned}&\left|\frac{(2i+2j-h)^2}{h} - \frac{(2i+2j-h-4)^2}{h}\right|\\ =\;& \left|\frac{4\{4(i+j)-2h-4\}}{h}\right|\\ \leq\;& \frac{4(2h-4)}{h} \leq \frac{4(4n-4)}{2n} = \frac{8(n-1)}{n}\end{aligned}$$

(ii) $(i,j) \leftrightarrow (i-2,j+1), (i-1,j), (i,j-1), (i+1,j-2)$

$$\begin{aligned}&\left|\frac{(2i+2j-h)^2}{h} - \frac{(2i+2j-h-2)^2}{h}\right|\\ =\;& \left|\frac{2\{4(i+j)-2h-2\}}{h}\right|\\ \leq\;& \frac{2(2h-2)}{h} \leq \frac{2(4n-2)}{2n} = \frac{2(2n-1)}{n}\end{aligned}$$

(iii) $(i,j) \leftrightarrow (i-2,j+2),(i-1,j+1),(i,j),(i+1,j-1),(i+2,j-2)$
$\chi^2_{hs}$ does not change.
(iv) $(i,j) \leftrightarrow (i-1,j+2),(i,j+1),(i+1,j),(i+2,j-1)$
Same as the case (ii).
(v) $(i,j) \leftrightarrow (i,j+2),(i+1,j+1),(i+2,j)$
Same as the case (i).

(II) When $h$ changes to $h+1$.

The change in $\chi^2_{hs}(h,i,j)$ is as follows:
(i) $(i,j) \leftrightarrow (i-1,j),(i,j-1)$

$$\begin{aligned} & \left| \frac{(2i+2j-h)^2}{h} - \frac{(2i+2j-h-3)^2}{h+1} \right| \\ = & \left| \frac{4(i+j)^2 + 8h(i+j) - 5h^2 - 9h}{h(h+1)} \right| \\ \leq & \frac{7h-9}{h+1} \leq \frac{7n-8}{n} \end{aligned}$$

(ii) $(i,j) \leftrightarrow (i-1,j+1),(i,j),(i+1,j-1)$

$$\begin{aligned} & \left| \frac{(2i+2j-h)^2}{h} - \frac{(2i+2j-h-1)^2}{h+1} \right| \\ = & \left| \frac{4(i+j)^2 - h^2 - h}{h(h+1)} \right| \leq \frac{3h-1}{h+1} \leq \frac{3n-2}{n} \end{aligned}$$

(iii) $(i,j) \leftrightarrow (i-1,j+2),(i,j+1),(i+1,j),(i+2,j-1)$

$$\begin{aligned} & \left| \frac{(2i+2j-h)^2}{h} - \frac{(2i+2j-h+1)^2}{h+1} \right| \\ = & \left| \frac{4(i+j)^2 - 8h(i+j) + 3h^2 - h}{h(h+1)} \right| \\ \leq & \frac{3h-1}{h+1} \leq \frac{3n-2}{n} \end{aligned}$$

(iv) $(i,j) \leftrightarrow (i,j+2),(i+1,j+1),(i+2,j)$

$$\begin{aligned} & \left| \frac{(2i+2j-h)^2}{h} - \frac{(2i+2j-h+3)^2}{h+1} \right| \\ = & \left| \frac{4(i+j)^2 - 16h(i+j) + 7h^2 - 9h}{h(h+1)} \right| \\ \leq & \frac{7h-9}{h+1} \leq \frac{7n-8}{n}. \end{aligned}$$

(III) When $h$ changes to $h+2$.

The change in $\chi^2_{hs}(h,i,j)$ is as follows:
(i) $(i,j) \leftrightarrow (i,j)$

$$\begin{aligned} & \left| \frac{(2i+2j-h)^2}{h} - \frac{(2i+2j-h-2)^2}{h+2} \right| \\ = & \left| \frac{8(i+j)^2 - 2h^2 - 4h}{h(h+2)} \right| \leq \frac{6h-4}{h+2} \leq \frac{2(3n-4)}{n} \end{aligned}$$

(ii) $(i,j) \leftrightarrow (i,j+1),(i+1,j)$

$$\begin{aligned} & \left| \frac{(2i+2j-h)^2}{h} - \frac{(2i+2j-h)^2}{h+2} \right| \\ = & \left| \frac{2(2i+2j-h)^2}{h(h+2)} \right| \leq \frac{2h}{h+2} \leq \frac{2(n-1)}{n} \end{aligned}$$

(iii) $(i,j) \leftrightarrow (i,j+2),(i+1,j+1),(i+2,j)$

$$\begin{aligned} & \left| \frac{(2i+2j-h)^2}{h} - \frac{(2i+2j-h+2)^2}{h+2} \right| \\ = & \left| \frac{8(i+j)^2 - 16h(i+j) + 6h^2 - 4h}{h(h+2)} \right| \\ \leq & \frac{6h-4}{h+2} \leq \frac{2(3n-4)}{n}. \end{aligned}$$

(IV) When $h$ changes to $h-1$ or $h-2$

Same as the case (II) and (III).
From the above, the sensitivity of $\chi^2_{hs}$ is $\frac{8(n-1)}{n}$. □

**Theorem S3.** *The sensitivity of the statistic $\chi^2_{total}$ for $n$ families with each family having two affected children is $\frac{16n-11}{n}$.*

*Proof.* The statistic $\chi^2_{total}$ can be expressed as a function

$$\chi^2_{total} : \mathcal{D} \longrightarrow \mathbb{R}_{\geq 0},$$

where $\mathcal{D} = \{(h,i,j) \in \mathbb{N} \mid 0 \leq i,j \leq h, i+j \leq h, 10 < h \leq 2n\}$.

As in the Theorem S1, we divide the cases according to the amount of change in the value of $h$.

(I) When the value of $h$ does not change.

The change in the value of $(i,j)$ are as follows:

$$(i,j) \leftrightarrow \begin{cases} (i-a,j+b) & (a \in \{1,2\}, b \in \{1,2\}) \\ (i \pm 1, j \pm 1) & (\text{double-sign corresponds}) \\ (i+a,j-b) & (a \in \{1,2\}, b \in \{1,2\}) \\ (i \pm 1, j),(i, j \pm b) & (a \in \{1,2\}, b \in \{1,2\}) \\ (i,j) & \end{cases}.$$

(i) $(i,j) \leftrightarrow (i-a,j+b)$

$$\begin{aligned} & h\left(\chi^2_{total}(h,i,j) - \chi^2_{total}(h,i-a,j+b)\right) \\ = & 4(3a-b)i + 4(a-3b)j \\ & -6a^2 + 4ab - 6b^2 - 4h(a-b) \end{aligned} \tag{5}$$

If $a \geq b$, i.e., $(a,b) = (1,1),(2,1),(2,2)$, then $a \leq i \leq h$, $0 \leq j \leq h-a$, $i+j \leq h$. Since $3a-b > 0$ and $a-3b < 0$,

$$\begin{aligned} (5) \geq & 4(3a-b)a + 4(a-3b)(h-a) \\ & -6a^2 + 4ab - 6b^2 - 4h(a-b) \\ = & 2a^2 + 12ab - 6b^2 - 8bh \geq 16(2-h), \end{aligned}$$

$$\begin{aligned} (5) \leq & 4(3a-b)h - 6a^2 + 4ab - 6b^2 - 4h(a-b) \\ = & -6a^2 + 4ab - 6b^2 + 8ah \leq 2(8h-11). \end{aligned}$$

Therefore,

$$\begin{aligned} & \left|\chi^2_{total}(h,i,j) - \chi^2_{total}(h,i-a,j+b)\right| \\ \leq & \frac{2(8h-11)}{h} \leq \frac{2(16n-11)}{2n} = \frac{16n-11}{n} \end{aligned}$$

If $a < b$, same as in the case of $a \geq b$.
(ii) $(i,j) \leftrightarrow (i \pm 1, j \pm 1)$

The maximum amount of change in $\chi^2_{total}$ corresponds to the sensitivity of $\chi^2_{hs}$, which is $\frac{8(n-1)}{n}$.

(iii) $(i,j) \leftrightarrow (i+a, j-b)$

Same as the case (i).

(iv) $(i,j) \leftrightarrow (i-a, j)$

$$\begin{aligned} & h\left(\chi^2_{total}(h,i,j) - \chi^2_{total}(h,i-a,j)\right) \\ = \;& 12ai + 4aj - 6a^2 - 4ah \end{aligned} \tag{6}$$

Since $a \le i \le h$, $0 \le j \le h-a$, and $i+j \le h$,

$$\begin{aligned} (6) \;&\ge\; 12a^2 - 6a^2 - 4ah \\ &=\; 6a^2 - 4ah \ge 24 - 8h, \end{aligned}$$

$$\begin{aligned} (6) \;&\le\; 4ah + 8ah - 6a^2 - 4ah \\ &=\; 8ah - 6a^2 \le 16h - 24. \end{aligned}$$

Therefore,

$$\begin{aligned} & \left|\chi^2_{total}(h,i,j) - \chi^2_{total}(h,i-a,j)\right| \\ \le \;& \frac{8(2h-3)}{h} \le \frac{8(4n-3)}{2n} = \frac{4(4n-3)}{n} \end{aligned}$$

(v) $(i,j) \leftrightarrow (i+a,j), (i, j \pm b)$

Same as in the case (iv).

(vi) $(i,j) \leftrightarrow (i,j)$

$\chi^2_{total}$ does not change.

(II) When $h$ changes to $h+1$.

(i) $(i,j) \leftrightarrow (i,j), (i \pm 1, j), (i, j \pm 1), (i-1, j+1), (i+1, j-1), (i+1, j+1)$

From the proofs of Theorem S1 and Theorem S2, the amount of change in $\chi^2_{total}$ is not more than

$$\frac{10n-9}{n} + \frac{3n-2}{n} = \frac{13n-11}{n}.$$

(ii) $(i,j) \leftrightarrow (i-1, j+2)$

$$\begin{aligned} & \chi^2_{total}(h,i,j) - \chi^2_{total}(h+1,i-1,j+2) \\ = \;& \frac{6i^2 + 4ij + 6j^2 - 24hi + 5h^2 - 22h}{h(h+1)} \end{aligned} \tag{7}$$

Since $0 \le i \le h$, $0 \le j \le h$, and $i+j \le h$,

$$(7) \;<\; \frac{11h^2 - 22h}{h(h+1)} = \frac{11(h-2)}{h+1},$$

$$(7) \;>\; \frac{-13h^2 - 22h}{h(h+1)} = \frac{-13h-22}{h+1}.$$

(iii) $(i,j) \leftrightarrow (i+2, j-1)$

Same as in the case (ii).

(iv) $(i,j) \leftrightarrow (i, j+2)$

$$\begin{aligned} & \chi^2_{total}(h,i,j) - \chi^2_{total}(h+1,i,j+2) \\ = \;& \frac{6i^2 + 4ij + 6j^2 - 12hi - 28hj + 9h^2 - 24h}{h(h+1)} \end{aligned} \tag{8}$$

Since $0 \le i \le h$, $0 \le j \le h$, and $i+j \le h$,

$$(8) \;<\; \frac{9h^2 - 24h}{h(h+1)} = \frac{3(3h-8)}{h+1},$$

$$(8) \;>\; \frac{-13h^2 - 24h}{h(h+1)} = \frac{-13h-24}{h+1}.$$

(v) $(i,j) \leftrightarrow (i+2, j)$

Same as in the case (iv).

(III) When $h$ changes to $h+2$.

(i) $(i,j) \leftrightarrow (i,j), (i, j+1), (i+1, j), (i+1, j+1)$

From the proofs of Theorem S1 and Theorem S2, the amount of change in $\chi^2_{total}$ is not more than

$$\frac{8n-9}{n} + \frac{2(n-1)}{n} = \frac{10n-11}{n}.$$

(ii) $(i,j) \leftrightarrow (i, j+2)$

$$\begin{aligned} & \chi^2_{total}(h,i,j) - \chi^2_{total}(h+2,i,j+2) \\ = \;& \frac{12i^2 + 8ij + 12j^2 - 24hi - 24hj + 10h^2 - 16h}{h(h+1)} \end{aligned} \tag{9}$$

Since $0 \le i \le h$, $0 \le j \le h$, and $i+j \le h$,

$$(9) \;\le\; \frac{10h^2 - 16h}{h(h+2)} = \frac{2(5h-8)}{h+2},$$

$$(9) \;>\; \frac{-6h^2 - 16h}{h(h+2)} = \frac{-2(3h+8)}{h+2}.$$

(iii) $(i,j) \leftrightarrow (i+2, j)$

Same as the case (ii).

(IV) When $h$ changes to $h-1$ or $h-2$.

Same as the case (II) and (III).

Consequently, the sensitivity of $\chi^2_{total}$ is $\frac{16n-11}{n}$. □

### B. Exponential Mechanism

There are ten possibilities shown in below for $(h,i,j)$ in one family to obtain these statistics, and we denote the number of families in each category as $n_1, n_2, \ldots, n_{10}$, in that order.

$$\begin{aligned} (h,i,j) \;=\; & (0,0,0), (1,0,0), (1,0,1), (1,1,0), (2,0,0), \\ & (2,0,1), (2,0,2), (2,1,0), (2,1,1), (2,2,0) \end{aligned}$$

#### 1) Exact Algorithm:

**Theorem S4.** *Algorithm 2 outputs the exact SHD score.*

*Proof.* We consider two cases: (I) $T < c^*$ and (II) $T \ge c^*$.

(I) $T < c^*$

To make $\chi^2_{td}$ larger, we need to increase the value of $|i-j|$.

First, we consider making $i$ larger than $j$. We start by looking at the case of changing one family in the category $(h,i,j) = (0,0,0)$. In this case, we can think about only three possible changes as follows: (i) $(0,0,0) \to (1,1,0)$, (ii)

$(0,0,0) \to (2,1,0)$, and (iii) $(0,0,0) \to (2,2,0)$. For each of these cases, the statistics after the change are given below:

$$\text{(i)}\ \frac{2(i-j+1)^2}{h+1},\ \text{(ii)}\ \frac{2(i-j+1)^2}{h+2},\ \text{(iii)}\ \frac{2(i-j+2)^2}{h+2}.$$

If $i > j$, the largest change is in the case (iii), so we change the family into the category $(2,2,0)$. When a family is in the categories $(1,0,0)$, $(1,0,1)$, $(1,1,0)$, $(2,0,0)$, $(2,0,1)$, $(2,0,2)$, $(2,1,0)$, and $(2,1,1)$, we can change it into the category $(2,2,0)$ as well.Then, since

$$\begin{aligned}&\frac{2(i-j+4)^2}{h} > \frac{2(i-j+3)^2}{h} > \frac{2(i-j+3)^2}{h+1}\\ &> \frac{2(i-j+2)^2}{h} > \frac{2(i-j+2)^2}{h+1} > \frac{2(i-j+2)^2}{h+2}\\ &> \frac{2(i-j+1)^2}{h} > \frac{2(i-j+1)^2}{h+1},\end{aligned}$$

we can check the number of categories as per Algorithm 2.

When making $i$ smaller than $j$, the proof is very similar to the above.

(II) $T \geq c^*$

When $i > j$, we can think as in the case (I) and change the families so that $\chi^2_{td}$ becomes smaller. In this case, we consider increasing the number of families included in the category $(2,0,2)$. Since

$$\begin{aligned}&\frac{2(i-j-4)^2}{h} < \frac{2(i-j-3)^2}{h+1} < \frac{2(i-j-3)^2}{h}\\ &< \frac{2(i-j-2)^2}{h+2} < \frac{2(i-j-2)^2}{h+1} < \frac{2(i-j-2)^2}{h}\\ &< \frac{2(i-j-1)^2}{h+1} < \frac{2(i-j-1)^2}{h},\end{aligned}$$

we can check the number of categories as per Algorithm 2.

When $i < j$, same as the case of $i > j$. □

**Theorem S5.** *Algorithm 3 outputs the exact SHD score.*

*Proof.* We consider two cases: (I) $T < c^*$ and (II) $T \geq c^*$.

(I) $T < c^*$

To make $\chi^2_{hs}$ larger, we need to increase the value of $|(i+j) - h/2|$.

First, we consider making $i+j$ larger than $h/2$. As in the case of Theorem S4, we can change a family into the category $(2,0,2)$, or $(2,1,1)$, or $(2,2,0)$. Here, since

$$\begin{aligned}&\frac{(2i+2j-h+4)^2}{h} > \frac{(2i+2j-h+3)^2}{h+1}\\ &> \frac{(2i+2j-h+2)^2}{h} > \frac{(2i+2j-h+2)^2}{h+2}\\ &> \frac{(2i+2j-h+1)^2}{h+1},\end{aligned}$$

we can check the number of categories as per Algorithm 3.

When making $i+j$ smaller than $h/2$, the proof is very similar to the above.

**Algorithm 2** Exact algorithm to find the SHD Score for $\chi^2_{td}$.

**Input:** Information about a single SNP, i.e., $n_k$ ($k = 1,\dots,10$), and the threshold $c^*$ for $\chi^2_{td}$.
**Output:** The SHD score in one SNP.

$h = \sum_{k=2}^{4} n_k + 2\sum_{k=5}^{10} n_k$
$i = n_4 + n_8 + n_9 + 2n_{10}$, $j = n_3 + n_6 + 2n_7 + n_9$
$T = 2(i-j)^2/h$

**if** $T < c^*$ **then**
  *Increase the number of families with* $(h,i,j) = (2,2,0)$.
  $d_1 = 0$, $N_k = n_k (k = 1,\dots,10)$
  **while** $T < c^*$ **do**
    Check the value of $N_7$, $N_6$, $N_3$, $N_5$, $N_9$, $N_2$, $N_1$, $N_8$, and $N_4$ in that order, and if a value greater than 0 is found, decrease it by 1 then continue to the next step.
    $N_{10} \leftarrow N_{10} + 1$
    $h = \sum_{k=2}^{4} N_k + 2\sum_{k=5}^{10} N_k$
    $i = N_4 + N_8 + N_9 + 2N_{10}$, $j = N_3 + N_6 + 2N_7 + N_9$
    $T = 2(i-j)^2/h$
    $d_1 \leftarrow d_1 - 1$
  **end while**

  *Increase the number of families with* $(h,i,j) = (2,0,2)$.
  $d_2 = 0$, $N_k = n_k (k = 1,\dots,10)$
  As in the above case, check $N_{10}$, $N_8$, $N_4$, $N_5$, $N_9$, $N_2$, $N_1$, $N_6$, and $N_3$ in that order, and increase $N_7$, then decrease $d_2$ until $T \geq c^*$.

  The SHD score is $\max\{d_1, d_2\}$.

**else if** $T \geq c^*$ **then**
  **if** $i > j$ **then**
    As in the case of $T < c^*$, check $n_{10}$, $n_4$, $n_8$, $n_1$, $n_2$, $n_5$, $n_9$, $n_3$, and $n_6$ in that order, and increase $n_7$ until $T < c^*$.
  **else**
    Check $n_7$, $n_3$, $n_6$, $n_1$, $n_2$, $n_5$, $n_9$, $n_4$, and $n_8$ in that order, and increase $n_{10}$ until $T < c^*$.
  **end if**
  The SHD score is (the number of steps) $-1$.
**end if**

(II) $T \geq c^*$

When $i + j > h/2$, we can think as in the case (I) and change the families so that $\chi^2_{hs}$ becomes smaller. In this case, we consider increasing the number of families included in the category $(2,0,2)$, or $(2,1,1)$, or $(2,2,0)$. Since

$$\begin{aligned}\frac{(2i+2j-h-4)^2}{h} &< \frac{(2i+2j-h-3)^2}{h+1}\\ &< \frac{(2i+2j-h-2)^2}{h+2} < \frac{(2i+2j-h-2)^2}{h}\\ &< \frac{(2i+2j-h-1)^2}{h+1},\end{aligned}$$

we can check the number of categories as per Algorithm 3.

When $i+j<h/2$, same as the case of $i+j>h/2$. □

**Algorithm 3** Exact algorithm to find the SHD Score for $\chi^2_{hs}$.

**Input:** Information about a single SNP, i.e., $n_k$ ($k = 1,\ldots,10$), and the threshold $c^*$ for $\chi^2_{hs}$.
**Output:** The SHD score in one SNP.

$h = \sum_{k=2}^{4} n_k + 2\sum_{k=5}^{10} n_k$
$i = n_4 + n_8 + n_9 + 2n_{10}$, $j = n_3 + n_6 + 2n_7 + n_9$
$T = (2i+2j-h)^2/h$

**if** $T < c^*$ **then**
  *Increase the number of families with* $(h,i,j) = (2,2,0)$.
  $d_1 = 0$, $N_k = n_k (k = 1,\ldots,10)$
  **while** $T < c^*$ **do**
    Check the value of $N_5$, $N_2$, $N_6$, $N_8$, $N_1$, $N_3$, and $N_4$ in that order, and if a value greater than 0 is found, decrease it by 1 then continue to the next step.
    $N_{10} \leftarrow N_{10} + 1$
    $h = \sum_{k=2}^{4} N_k + 2\sum_{k=5}^{10} N_k$
    $i = N_4 + N_8 + N_9 + 2N_{10}$, $j = N_3 + N_6 + 2N_7 + N_9$
    $T = (2i+2j-h)^2/h$
    $d_1 \leftarrow d_1 - 1$
  **end while**

  *Increase the number of families with* $(h,i,j) = (2,0,0)$.
  $d_2 = 0$, $N_k = n_k (k = 1,\ldots,10)$
  As in the above case, check $N_7$, $N_9$, $N_{10}$, $N_3$, $N_4$, $N_6$, $N_8$, $N_1$, and $N_2$ in that order, and increase $N_5$, then decrease $d_2$ until $T \geq c^*$.

  The SHD score is $\max\{d_1, d_2\}$.

**else if** $T \geq c^*$ **then**
  **if** $i+j > h/2$ **then**
    As in the case of $T < c^*$, check $n_7$, $n_9$, $n_{10}$, $n_3$, $n_4$, $n_1$, $n_6$, $n_8$, and $n_2$ in that order, and increase $n_5$ until $T < c^*$.
  **else**
    Check $n_5$, $n_2$, $n_1$, $n_6$, $n_8$, $n_3$, and $n_4$ in that order, and increase $n_{10}$ until $T < c^*$.
  **end if**
  The SHD score is (the number of steps) $-1$.
**end if**

*2) Approximation Algorithm:*

**Theorem S6.** *The sensitivity of the SHD score obtained by Algorithm 5 is 1.*

*Proof.*

(I) $2(i-j)^2/h < c^*$

(i) $h \leq \frac{c^*}{2}$

When the changes in $h$ are 2, 1, and 0, the maximum changes in $|i-j|$ are 2, 3, and 4, respectively. Therefore, the SHD score changes by at most $\lceil \frac{4}{4} \rceil = 1$.

(ii) $h > \frac{c^*}{2}$

When the change in $h$ is 2, the maximum change in $|i-j|$ is 2. Therefore, we can consider the following inequality:

$$\begin{aligned}&\left(\sqrt{\frac{(h+2)c^*}{2}} - (|i-j|-2)\right) - \left(\sqrt{\frac{(hc^*}{2}} - (|i-j|)\right)\\ &= \sqrt{\frac{(h+2)c^*}{2}} - \sqrt{\frac{hc^*}{2}} + 2 = \sqrt{\frac{c^*}{2}}\left(\sqrt{h+2} - \sqrt{h}\right) + 2\\ &= \sqrt{\frac{c^*}{2}} \cdot \frac{2}{\sqrt{h+2}+\sqrt{h}} + 2\\ &\leq \sqrt{\frac{c^*}{2}} \cdot \frac{2}{2\sqrt{h}} + 2\\ &< 3 \quad [\because h > \frac{c^*}{2}]\end{aligned}$$

Similarly, when the changes in $h$ are 1 and 0, $\sqrt{\frac{hc^*}{2}} - |i-j|$ changes by at most $\frac{7}{2}$ and 4, respectively.

Therefore, the SHD score changes by at most $\lceil \frac{4}{4} \rceil = 1$.

(II) $2(i-j)^2/h \geq c^*$

Same as the case (I)(ii). □

**Algorithm 5** Approximation algorithm to find the SHD Score for $\chi^2_{td}$.

**Input:** Information about a single SNP, i.e., $n_k$ ($k = 1,\ldots,10$), and the threshold $c^*$ for $\chi^2_{td}$.
**Output:** The SHD score in one SNP.

$h = \sum_{k=2}^{4} n_k + 2\sum_{k=5}^{10} n_k$
$i = n_4 + n_8 + n_9 + 2n_{10}$, $j = n_3 + n_6 + 2n_7 + n_9$
$\chi^2_{td} = 2(i-j)^2/h$
**if** $\chi^2_{td} < c^*$ **then**
  **if** $h \leq c^*/2$ **then**
    The SHD score is $-\left\lceil \frac{c^*-h-|i-j|}{4} \right\rceil$.
  **else if** $h > c^*/2$ **then**
    The SHD score is $-\left\lceil \frac{\sqrt{hc^*/2}-|i-j|}{4} \right\rceil$.
  **end if**
**else if** $\chi^2_{td} \geq c^*$ **then**
  The SHD score is $\left\lceil \frac{|i-j|-\sqrt{hc^*/2}}{4} \right\rceil - 1$.
**end if**

**Theorem S7.** *The sensitivity of the SHD score obtained by Algorithm 6 is 1.*

*Proof.*

(I) $(2i+2j-h)^2/h < c^*$

(i) $i+j \geq h/2$

For the case of $h \leq c^*$, it is enough to note that the maximum change in $i+j$ is 2. In the following, we consider the case of $h > c^*$.

When the change in $h$ is 2, the maximum change in $-(i+j)$ is 0. Since

$$\begin{aligned}\sqrt{(h+2)c^*} - \sqrt{hc^*} &= \sqrt{c^*} \cdot \frac{2}{\sqrt{h+2}+\sqrt{h}} \\ &< \sqrt{c^*} \cdot \frac{1}{\sqrt{h}} < 1, \quad [\because h > c^*]\end{aligned}$$

the SHD score changes at most

$$\left\lceil \frac{\frac{2+1}{2}+0}{2} \right\rceil = \left\lceil \frac{3}{4} \right\rceil = 1.$$

Similarly, when the changes in $h$ are 1 and 0, the SHD score changes by at most $\left\lceil \frac{7}{8} \right\rceil = \lceil 1 \rceil = 1$.

(ii) $i+j < h/2$

For the case of $h > c^*$, as in the case (i), the maximum change in the SHD score is 1. Then, we consider the case of $h \leq c^*$. When the changes in $h$ are 2, 1, and 0, $-(i+j)$ changes by at most 0, 1, and 2, respectively. Therefore, the SHD score changes by at most $\left\lceil \frac{2}{2} \right\rceil = 1$.

(II) $2(i-j)^2/h \geq c^*$

Similar to the case (I). □

**Theorem S8.** *The sensitivity of the SHD score obtained by Algorithm 7 is 1.*

*Proof.*

(I) $h \leq c^*$

$\frac{dc^*}{2\sqrt{2}}$ is the shortest distance between the ellipse $\chi^2_{total} = c^*$ and the point $(i,j)$ when $h$ is fixed to $c^*$.

When the change in $h$ is 2, the maximum distance moved by the point $(i,j)$ is 2. Therefore, the SHD score changes by at most

$$\left\lceil \frac{(\sqrt{2}-1)\cdot 2}{2\sqrt{2}} + \frac{2}{2\sqrt{2}} \right\rceil = 1.$$

Similarly, when the changes in $h$ are 1 and 0, the maximum changes in the SHD score are

$$\left\lceil \frac{(\sqrt{2}-1)\cdot 1}{2\sqrt{2}} + \frac{\sqrt{5}}{2\sqrt{2}} \right\rceil = 1, \quad \left\lceil \frac{2\sqrt{2}}{2\sqrt{2}} \right\rceil = 1,$$

respectively.

(II) $h > c^*$

$\frac{d\sqrt{hc^*}}{2\sqrt{2}}$ is the shortest distance between the ellipse $\chi^2_{total} = c^*$ and the point $(i,j)$.

**Algorithm 6** Approximation algorithm to find the SHD Score for $\chi^2_{hs}$.

**Input:** Information about a single SNP, i.e., $n_k$ ($k = 1,\ldots,10$), and the threshold $c^*$ for $\chi^2_{hs}$.
**Output:** The SHD score in one SNP.

$h = \sum_{k=2}^{4} n_k + 2\sum_{k=5}^{10} n_k$
$i = n_4 + n_8 + n_9 + 2n_{10}$, $j = n_3 + n_6 + 2n_7 + n_9$
$\chi^2_{hs} = (2i+2j-h)^2/h$
**if** $\chi^2_{hs} < c^*$ **then**
  **if** $i+j \geq h/2$ **then**
    **if** $h \leq c^*$ **then**
      The SHD score is $-\left\lceil \frac{c^*-(i+j)}{2} \right\rceil$.
    **else if** $h > c^*$ **then**
      The SHD score is $-\left\lceil \frac{(h+\sqrt{hc^*})/2-(i+j)}{2} \right\rceil$.
    **end if**
  **else if** $i+j < h/2$ **then**
    **if** $h \leq c^*$ **then**
      The SHD score is $-\left\lceil \frac{c^*-h+(i+j)}{2} \right\rceil$.
    **else if** $h > c^*$ **then**
      The SHD score is $-\left\lceil \frac{(i+j)-(h-\sqrt{hc^*})/2}{2} \right\rceil$.
    **end if**
  **end if**
**else if** $\chi^2_{hs} \geq c^*$ **then**
  **if** $i+j \geq h/2$ **then**
    The SHD score is $\left\lceil \frac{(i+j)-(h+\sqrt{hc^*})/2}{2} \right\rceil - 1$.
  **else if** $i+j < h/2$ **then**
    The SHD score is $\left\lceil \frac{(h-\sqrt{hc^*})/2-(i+j)}{2} \right\rceil - 1$.
  **end if**
**end if**

When the change in $h$ is 2, the distance moved by the center of the ellipse is $\frac{\sqrt{2}}{2}$. The semi-major and semi-minor axes can be increased by at most $\frac{1}{2}$ and $\frac{\sqrt{2}}{4}$, respectively, and the maximum distance moved by the point $(i,j)$ is 2. By considering the direction of each change, the shortest distance between the ellipse and the point $(i,j)$ varies by at most 2.

Similarly, when the changes in $h$ are 1 and 0, the maximum change in the shortest distance are $\sqrt{5}$ and $2\sqrt{2}$, respectively.

Therefore, the SHD score changes by at most $\left\lceil \frac{2\sqrt{2}}{2\sqrt{2}} \right\rceil = 1$. □

## S3. Experiments

### A. Simulation Data

For both cases in (I) small cohort and (II) large cohort, we generated simulation data for $\chi^2_{td}$, $\chi^2_{hs}$, and $\chi^2_{total}$. We show the generation procedures and the distributions of the statistics in the datasets below.

(I) Small Cohort

We set the family number $N = 150$ and SNP number $M = 5,000$ as in the experiments by Wang et al. [5], and

**Algorithm 7** Approximation algorithm to find the SHD Score for $\chi^2_{total}$.

**Input:** Information about a single SNP, i.e., $n_k$ ($k = 1, \dots, 10$), and the threshold $c^*$ for $\chi^2_{total}$.
**Output:** The SHD score in one SNP.

$h = \sum_{k=2}^{4} n_k + 2\sum_{k=5}^{10} n_k$
$i = n_4 + n_8 + n_9 + 2n_{10}$, $j = n_3 + n_6 + 2n_7 + n_9$
$\chi^2_{total} = \{4(i-h/4)^2 + 2(i+j-h/2)^2 + 4(j-h/4)^2\}/h$
**if** $h \leq c^*$ **then**
  $s = 2(i-j)/c^*$, $t = 2(i+j-c^*/2)/c^*$
  Let $d$ be the shortest distance between the point $(s,t)$ and the ellipse $\frac{x^2}{2} + y^2 = 1$.
  **if** $\chi^2_{total} < c^*$ **then**
    The SHD score is $-\left\lceil \frac{(\sqrt{2}-1)(c^*-h)}{2\sqrt{2}} + \frac{dc^*}{8} \right\rceil$.
  **else if** $\chi^2_{total} \geq c^*$ **then**
    The SHD score is $\left\lceil \frac{(\sqrt{2}-1)(c^*-h)}{2\sqrt{2}} + \frac{dc^*}{8} \right\rceil - 1$.
  **end if**
**else if** $h > c^*$ **then**
  $s = 2(i-j)/\sqrt{hc^*}$, $t = 2(i+j-h/2)/\sqrt{hc^*}$
  Let $d$ be the shortest distance between the point $(s,t)$ and the ellipse $\frac{x^2}{2} + y^2 = 1$.
  **if** $\chi^2_{total} < c^*$ **then**
    The SHD score is $-\left\lceil \frac{d\sqrt{hc^*}}{8} \right\rceil$.
  **else if** $\chi^2_{total} \geq c^*$ **then**
    The SHD score is $\left\lceil \frac{d\sqrt{hc^*}}{8} \right\rceil - 1$.
  **end if**
**end if**

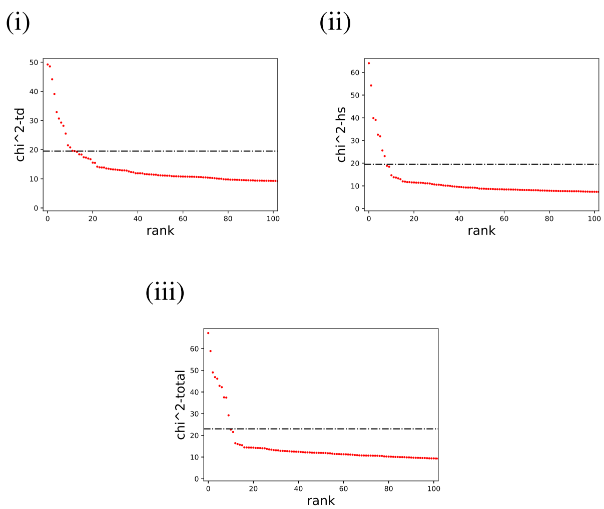


Fig. S1. The top 100 statistics in the simulation data for (i) $\chi^2_{td}$, (ii) $\chi^2_{hs}$, and (iii) $\chi^2_{total}$ in a small cohort. The dotted lines are thresholds at $100(1-0.05/M)$%-quantile of $\chi^2$-distribution based on the Bonferroni correction.

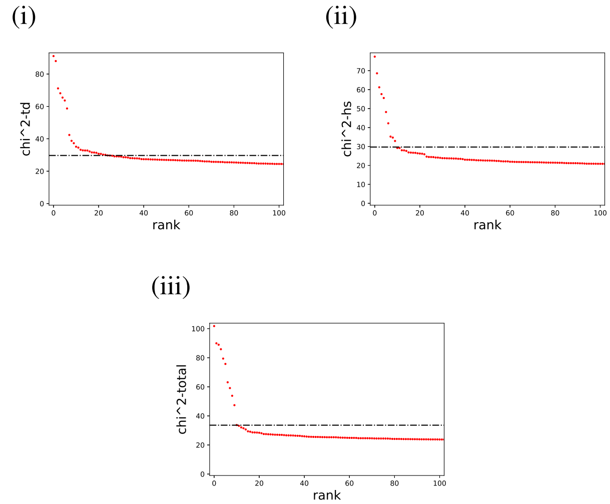


Fig. S2. The top 100 statistics in the simulation data for (i) $\chi^2_{td}$, (ii) $\chi^2_{hs}$, and (iii) $\chi^2_{total}$ in a large cohort. The dotted lines are thresholds at $100(1-0.05/M)$%-quantile of $\chi^2$-distribution based on the Bonferroni correction.

assume that about 10 SNPs are significant. Here, we consider generating a dataset for the $i$-th SNP.

(i) $\chi^2_{td}$

We set $n_1$ to $n_{10}$ by the following equations:

$$n_1 = Binomial\left(2N, \frac{1}{10}\right),$$
$$n_j = Binomial\left(2N - \sum_{k}^{j-1} n_k, \frac{1}{11-j}\right) (j = 2, \dots, 10).$$

For the generation of 10 significant datasets, when calculating $n_3$, $n_6$, and $n_7$, we multiply the probability in the binomial distribution by 1.14.

(ii) $\chi^2_{hs}$

We obtain $n_1$ to $n_{10}$ as follows:

$$n_5 = Binomial\left(2N, \frac{1}{5}\right),$$
$$S_i = Binomial\left(2N - n_5, \frac{1}{4}\right), n_7 = Binomial\left(S_i, \frac{1}{3}\right),$$
$$n_9 = Binomial\left(S_i - n_7, \frac{1}{2}\right), n_{10} = S_i - n_7 - n_9,$$
$$n_2 = Binomial\left(2N - n_5 - S_i, \frac{1}{3}\right),$$
$$T_i = Binomial\left(2N - n_5 - S_i - n_2, \frac{1}{2}\right),$$
$$n_3 = Binomial\left(T_i, \frac{1}{2}\right), n_4 = T_i - n_3,$$
$$R_i = 2N - n_5 - S_i - n_2 - T_i, n_1 = Binomial\left(R_i, \frac{1}{3}\right),$$
$$n_6 = Binomial\left(R_i - n_1, \frac{1}{2}\right), n_8 = R_i - n_1 - n_6.$$

For the generation of 10 significant datasets, when calculating $n_5$, and $n_2$, we multiply the probability in the binomial distribution by 1.1.

(iii) $\chi^2_{total}$

We obtain $n_1$ to $n_{10}$ as follows:

$$n_5 = Binomial\left(2N, \frac{1}{5}\right),$$

$$S_i = Binomial\left(2N - n_5, \frac{1}{4}\right), P_i = Binomial\left(S_i, \frac{1}{2}\right),$$

$$n_7 = Binomial\left(P_i, \frac{1}{2}\right), n_{10} = P_i - n_7, n_9 = S_i - P_i,$$

$$n_2 = Binomial\left(2N - n_5 - S_i, \frac{1}{3}\right),$$

$$T_i = Binomial\left(2N - n_5 - S_i - n_2, \frac{1}{2}\right),$$

$$n_3 = Binomial\left(T_i, \frac{1}{2}\right), n_4 = T_i - n_3,$$

$$R_i = 2N - n_5 - S_i - n_2 - T_i, Q_i = Binomial\left(R_i, \frac{1}{2}\right),$$

$$n_6 = Binomial\left(Q_i, \frac{1}{2}\right), n_8 = Q_i - n_6, n_1 = R_i - Q_i.$$

For the generation of 10 significant datasets, when calculating $n_5$, $n_7$, $n_2$, $n_3$, and $n_6$, we multiply the probability in the binomial distribution by 1.1.

(II) Large Cohort

We set $N = 5,000$ and $M = 10^6$ as in the experiments by Wang et al. [5] The way to generate non-significant datasets is the same as in (I). When generating 10 significant datasets, we multiply the probabilities in the corresponding binomial distribution by 1.6 for $\chi^2_{td}$ and by 1.5 for $\chi^2_{hs}$ and $\chi^2_{total}$.

The distributions of the statistics in datasets for a small cohort and a large cohort generated y the above procedure are shown in Fig. S1 and Fig. S2, respectively.

## References


[1] C. Dwork, "Differential privacy," In: M. Bugliesi, B. Preneel, V. Sassone, I. Wegener (eds) Automata, Languages and Programming, ICALP 2006, pp. 1–12, 2006.

[2] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," In: S. Halevi, T. Rabin (eds) Theory of Cryptography, TCC 2006, LNCS, vol. 3876, pp. 265–284, 2006.

[3] F. McSherry and K. Talwar, "Mechanism design via differential privacy," 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), pp. 94–103, October 2007.

[4] A. Johnson and V. Shmatikov, "Privacy-preserving data exploration in genome-wide association studies," In: KDD 2013, pp. 1079–1087, August 2013.

[5] M. Wang, Z. Ji, S. Wang, J. Kim, H. Yang, X. Jiang, and L. Ohno-Machado, "Mechanisms to protect the privacy of families when using the transmission disequilibrium test in genome-wide association studies," Bioinformatics, vol. 33, no. 23, pp. 3716–3725, December 2017.