



**HAL**  
open science

## Dépistages, terrorisme, et statistiques

David Monniaux

► **To cite this version:**

| David Monniaux. Dépistages, terrorisme, et statistiques. 2025. hal-04937842

**HAL Id: hal-04937842**

**<https://hal.science/hal-04937842v1>**

Submitted on 10 Feb 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Dépistages, terrorisme, et statistiques

David Monniaux, 10 février 2025

Régulièrement, des entrepreneurs promeuvent des solutions technologiques destinées à détecter automatiquement des personnes susceptibles de présenter tel ou tel risque (terroristes...). Rappelons quelques faits statistiques...

Imaginons que l'on passe un détecteur automatique de « profils à risque terroriste » sur les échanges Internet de 30 millions de résidents français, et que ce détecteur soit réputé avoir un taux d'exactitude de 99,9 %, ce qui peut paraître excellent à première vue. Supposons que cette prétention de l'entrepreneur technologique soit bien fondée (l'entrepreneur est honnête et compétent lorsqu'il énonce cela, ce qui est une hypothèse forte) et qu'elle signifie que ce détecteur a un taux de 0,1 % de faux positifs (0,1 % des non terroristes vont être détectés comme terroristes) et un taux de 0,1 % de faux négatifs (0,1 % des terroristes ne vont pas être détectés comme tels).

Supposons qu'il y ait 1000 profils terroristes parmi les 30 millions de personnes « scannées ». Parmi ces 1000, en moyenne, 999 seront détectées positives, et 1 ne le sera pas. Parmi les 30 millions moins 1000 profils non terroristes, en moyenne 29999, soit environ 30000, seront détectées comme terroristes potentiels. Ainsi, on aura environ 31000 détections positives, dont seulement 1000 seront des vraies pistes : le nombre de fausses pistes sera 30 fois supérieur à celui de vraies pistes ! Pareil dispositif, en apparence d'excellente précision, aura donc tendance à noyer les enquêteurs sous des flots d'indications qu'ils n'auront pas la possibilité de toutes suivre.

Ce problème, dit d'*oubli de la fréquence de base*, touche de nombreux domaines d'activité, indépendamment des technologies mises en œuvre, à chaque fois qu'on veut rechercher un phénomène rare. J'ai eu affaire à lui vers 2002 dans le contexte d'outils destinés à (grossièrement) rechercher automatiquement des bugs dans des programmes informatiques : lorsqu'on les applique à des logiciels où les bugs sont à la base rares (en l'occurrence, des commandes de vol d'avions de ligne, soit probablement parmi les logiciels les mieux contrôlés qui existent), la très grande majorité des rapports produits sont des faux positifs, et ce même si la publicité des outils indique de faibles taux d'erreur. Le même problème se pose, si j'ai bien compris ce que disent les médecins, pour les dépistages de maladies relativement rares : même avec de faibles taux d'erreurs, une grande partie des personnes diagnostiquées positives ne sont en réalité pas touchées par la maladie, alors qu'on les soumet à l'anxiété et à la prescription d'examen complémentaires, parfois invasifs.

La réponse à ce problème, dans le cas médical (toujours si j'ai bien compris ce que disent les médecins) est de cibler des groupes à risque, de sorte que la maladie, rare dans la population générale, est alors nettement moins rare. Il est possible que la même méthodologie s'applique dans le cas de la détection du terrorisme. Par ailleurs, tout ceci est bien évidemment relatif à des choix politiques, notamment le taux d'erreurs que l'on se permet — par exemple, [il semble qu'Israël se permette un certain taux de victimes civiles dans son ciblage des personnalités du Hamas](#).

Que faut-il en conclure ? Les arguments statistiques sont souvent délicats à comprendre. Des chiffres en apparence très encourageants peuvent en réalité être catastrophiques. La compréhension des probabilités conditionnelles est indispensable avant de se fier à quelque indication statistique que ce soit.

PS : Voir par exemple les [\*Éléments d'analyse technique du projet de loi relatif au renseignement\*](#) (INRIA, 30 avril 2015)