

# Échantillonnage de l'enquête Trajectoires et Origines 2

## Journées de méthodologie statistique de l'Insee 2022

Nicolas Paliod, Willy Thao Khamsing, Thomas Merly-Alpa

Insee - Ined

31 mars 2022

- 1 Présentation du dispositif d'enquête
- 2 Évolutions de l'échantillonnage des G1
- 3 Échantillonnage des G2
- 4 Conclusion

# Plan

- 1 Présentation du dispositif d'enquête
- 2 Évolutions de l'échantillonnage des G1
- 3 Échantillonnage des G2
- 4 Conclusion

# Objectifs de l'enquête

- TeO2 est la deuxième édition de l'enquête TeO menée en 2008-2009
- Un dispositif national unique ayant pour objectifs
  - de mesurer l'impact des origines sur les trajectoires sociales, les discriminations et l'intégration
  - d'évaluer la persistance de ces analyses sur plusieurs générations
- 2 populations d'intérêt : les immigrés et les descendants d'immigrés

## Mise en place de l'enquête TeO2

- Une co-maîtrise d'ouvrage Ined - Insee
- Un peu plus de 50 000 individus échantillonnés
- Répartition en 5 sous-échantillons
  - Un sous-échantillon d'immigrés (G1)
  - Un sous-échantillon d'individus nés dans les DROM (G1)
  - Un sous-échantillon d'individus dont les parents sont immigrés (G2)
  - Un sous-échantillon d'individus dont les parents sont nés dans les DROM (G2)
  - Un sous-échantillon en population générale

## Le dispositif TeO2 - G3, une nouveauté

- Une nouveauté de l'enquête TeO2 : l'intérêt pour la persistance des inégalités, les discriminations, etc. jusqu'à la 3<sup>e</sup> génération
- Mais les petits enfants d'immigrés d'origine non européenne sont une population très minoritaire et difficile à atteindre
- Choix de procéder à un sondage indirect pour suréchantillonner les petits enfants d'immigrés d'origine non européenne en demandant aux G2 répondants les coordonnées de leurs enfants
- Enquête à part et dispositif expérimental

# La suite de la présentation

## Axes de la présentation

- Présentation des évolutions par rapport à TeO pour l'échantillonnage des G1
- Présentation du dispositif complexe de l'échantillonnage des G2

## Points non abordés

- L'échantillonnage des G3
- L'échantillonnage de l'échantillon en population générale

# Plan

- 1 Présentation du dispositif d'enquête
- 2 Évolutions de l'échantillonnage des G1
- 3 Échantillonnage des G2
- 4 Conclusion



## Groupes d'intérêt de l'enquête TeO

- Algérie
- Maroc, Tunisie
- Afrique Sahélienne
- Afrique centrale et du golfe de Guinée
- Espagne, Italie, Portugal
- Autres pays de l'Union Européenne
- Turquie
- Asie du Sud-Est
- Autres pays du monde
- DROM

# Nouveautés de l'enquête TeO2

- 2 nouvelles populations d'intérêt
  - Individus immigrés de nationalité chinoise à la naissance
  - Individus immigrés depuis des pays dont proviennent de nombreux réfugiés

# Plan

- 1 Présentation du dispositif d'enquête
- 2 Évolutions de l'échantillonnage des G1
- 3 Échantillonnage des G2**
- 4 Conclusion

## Problématiques de l'échantillonnage des G2

- Interroger des individus dont les parents sont immigrés
- Aucune base de sondage ne dispose d'information sur le statut d'immigration des parents
- Seule solution : aller en mairie pour connaître le lieu de naissance des parents

## Groupes de population cible

- Algérie
- Maroc, Tunisie
- Afrique Sahélienne
- Afrique centrale et du golfe de Guinée
- Espagne, Italie
- Portugal
- Autres pays de l'Union Européenne
- Turquie
- Asie du Sud-Est
- Autres pays du monde
- DROM

## Processus d'échantillonnage (1/2)

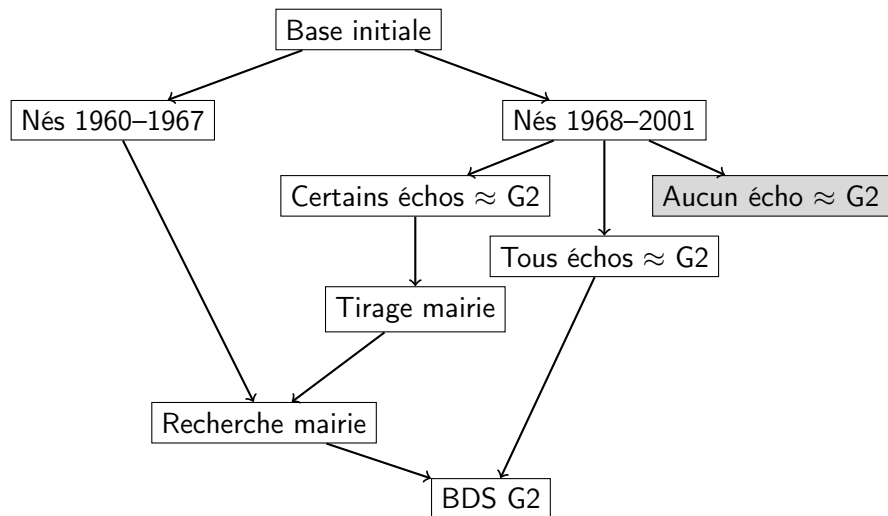
- Tirage dans l'enquête annuelle de recensement (EAR) 2018
  - Avantage de l'EAR : on connaît la nationalité de l'individu et on peut se restreindre aux individus de nationalité française à la naissance
- Appariement de l'EAR 2018 avec l'état civil anonymisé (Statec)
  - Avantage : on récupère des informations sur le lieu de naissance des parents
  - Désavantage : l'appariement se fait sur le sexe, la date de naissance et la commune de naissance de l'individu, donc à un individu de l'EAR peuvent correspondre plusieurs individus de Statec
  - Désavantage : Statec n'est disponible qu'à partir de 1968, aucun appariement possible pour les individus du champ nés en 1967 ou avant

## Processus d'échantillonnage (2/2)

### Solution retenue

- Sélection d'un échantillon de 100 000 individus pour lesquels il y a une incertitude sur le lieu de naissance des parents
  - Individus ayant plusieurs « échos » dans Statec, avec des surreprésentations en fonction du lieu de naissance potentiel des parents
  - Individus nés avant 1967 sans surreprésentation
- Vérification en mairie du lieu de naissance des parents des 100 000 individus tirés
- Les individus dont les parents sont effectivement nés à l'étranger ( $\neq$  immigrés) ou dans les DOM, soit avec certitude d'après Statec, soit d'après les relevés mairie, forment la base de sondage des G2

## Illustration





# Des poids très dispersés dans la base de sondage des G2 (1/2)

- De nombreuses étapes de tirage
  - Sélection liée à l'utilisation de l'EAR
  - Restriction à des unités primaires dont les poids sont proportionnels à la taille
  - Restriction à des jours de naissance plus ou moins nombreux en fonction de l'origine potentielle des parents
  - Dispersion des poids liée au tirage de relevés mairie (probabilité de tirage différente si l'individu est né en 1967 ou avant ou s'il est né en 1968 ou après)

## Des poids très dispersés dans la base de sondage des G2 (2/2)

- Un aléa lié à la collecte en mairie
  - Un individu appartenant à une strate surreprésentée pour le tirage des relevés mairie peut avoir des parents d'une origine non surreprésentée
- Poids de sondage très dispersés à l'issue des relevés mairie
- Solution :
  - Échantillonner les individus avec une probabilité proportionnelle à leur poids dans la base de sondage pour limiter la dispersion des poids au sein d'une strate donnée

## Des poids encore dispersés dans l'échantillon des G2

- Choix de sous-représenter l'Île-de-France : impossible de disposer de poids non dispersés au sein d'une strate
- Une dispersion résiduelle non négligeable des poids vient de la taille de la base de sondage issue des relevés mairie qui n'est pas assez grande par rapport à la taille d'échantillon : risque de valeurs influentes
- Exemples (strates extrêmes)

Strate de tirage / Distribution	Min	D1	D2	D3	D4	D5	D6	D7	D8	D9	Max
Afrique centrale et guinéenne	9	10	15	28	31	33	40	56	83	138	807
Italie Espagne	434	434	434	434	434	434	434	434	434	551	1338

# Plan

- 1 Présentation du dispositif d'enquête
- 2 Évolutions de l'échantillonnage des G1
- 3 Échantillonnage des G2
- 4 Conclusion**

# Conclusion

- Une enquête unique en France
- Quelques évolutions par rapport à TeO
- Un dispositif complet mais complexe
- La complexité du dispositif nécessite d'être vigilant au moment des redressements (partage des poids, valeurs influentes...)