

# Analyse démographique pour la prise des décisions en Afrique francophone : formation sur les outils en ligne de l'UNFPA / UIESP.

UIESP / UNFPA ATELIER POUR L'AFRIQUE FRANCOPHONE

*Dakar, Sénégal, 2-6 novembre 2015*

**Cheikh Tidiane NDIAYE**

Agence Nationale de la Statistique et de la Démographie  
(Sénégal)

**Fiabilité des données des recensements.  
Comparaisons avec d'autres sources pour mieux  
comprendre les erreurs et les biais**

---

# Plan

1. Pourquoi évaluer les données de recensement
  2. Quels types de tests ?
  3. Recours aux méthodes démographiques
    - a) Analyses des distributions par sexe et âge
    - b) Comparaison de recensements succesifs et données actuelles
    - c) Analyse des taux de survie par cohorte
    - d) Quelques exemples
  4. Recours aux méthodes d'appariement
    - a) Evaluer avec les données d'observatoires de population
    - b) Evaluer par une enquête post censitaire (EPC)
  5. Conclusion
-

# Pourquoi évaluer les données ?

- RGPH, une opération lourde
  - Problèmes de coopération des personnes.
  - Données affectées par des erreurs d'observation
    - Mode de collecte (enquêteurs, enquêtés, accès au terrain)
    - Mode de compilation (saisie, application électronique)
    - Omissions d'individus ou d'événements ;
    - doubles comptes ;
    - erreurs sur les périodes de référence (naissances, décès, etc. des 12 mois ayant précédé l'opération) ;
    - erreurs sur les âges, etc.
  - Identification et mesure de leur ampleur des erreurs en vue de :
    - Juger de la validité des données pour fournir une réponse d'une précision acceptable
    - Proposer des ajustements
-

# Quels types de tests ?

1. Test de cohérence (interne et externe)
  2. Comparaison de statistiques observées avec celles théoriquement attendues (cas population stable)
  3. Comparaison des statistiques avec celles d'une autre population socio-démographiquement similaire
  4. Comparaison avec d'autres sources de données (administratives, état civil, observatoires, etc.)
  5. Vérifications directes (ré-enquête comme EPC)
-

# Quand utiliser les méthodes d'évaluation ?

Estimer et évaluer la qualité des données

- Identifier la nature des imperfections
- Estimer leur ampleur
- Évaluer leur signification
- Comprendre les conséquences sur l'usage des données



# Sources

Les principes de bases ont peu changé

- Manuel II des Nations Unies : UN Population Branch, 1955.  
*Evaluation de la qualité des statistiques de base pour les estimations de la population*
-

# Analyse des distributions par sexe et âge

## A partir d'une seule source : le recensement

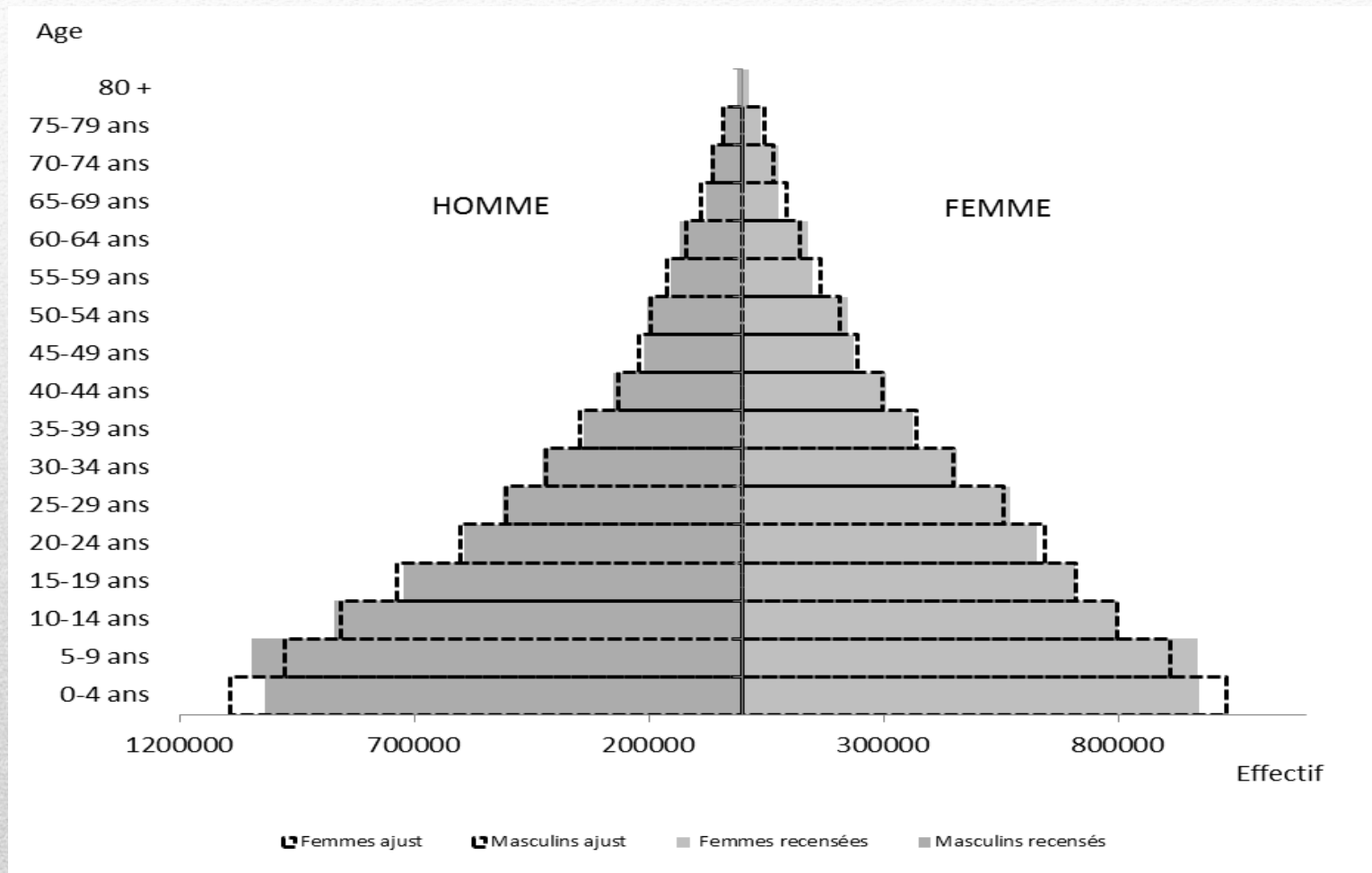
- Examiner la forme des distributions par sexe/âge de la population recensée
  - Rapprocher ces distributions (sexe/âge) avec la structure de la fécondité, mortalité et migration internationale
  - Comparer ces distributions avec celles attendues pour apprécier l'ampleur des écarts (erreurs dans le recensement)
  - Limites : difficulté à en déduire des estimations directes des taux de couverture.
  - Nécessite souvent de recourir à d'autres méthodes, comme l'EPC pour vérifier
  - Il est aussi difficile d'expliquer l'origine des incohérences entre les distributions attendues et observées
-

# Analyse des distributions par sexe et âge

- Cohérence interne
    - Ratios et indices d'écart absolus entre les distributions attendues (voire autre source) et observées
      - Représentations graphiques (pyramide; analyse de cohorte)
      - Ratios par âge/sexe
      - Statut matrimonial, etc.
    - Identification des âges attractifs (indices résumés)
      - Indices de Whipples et composite Myers (préférence pour les âges ronds)
      - Indice de précision âge-sex (Age-sex accuracy index)
-



# Analyse des distributions par sexe et âge



# Analyse des distributions par sexe et âge

Sources possibles des distorsions dans la pyramide des âges :

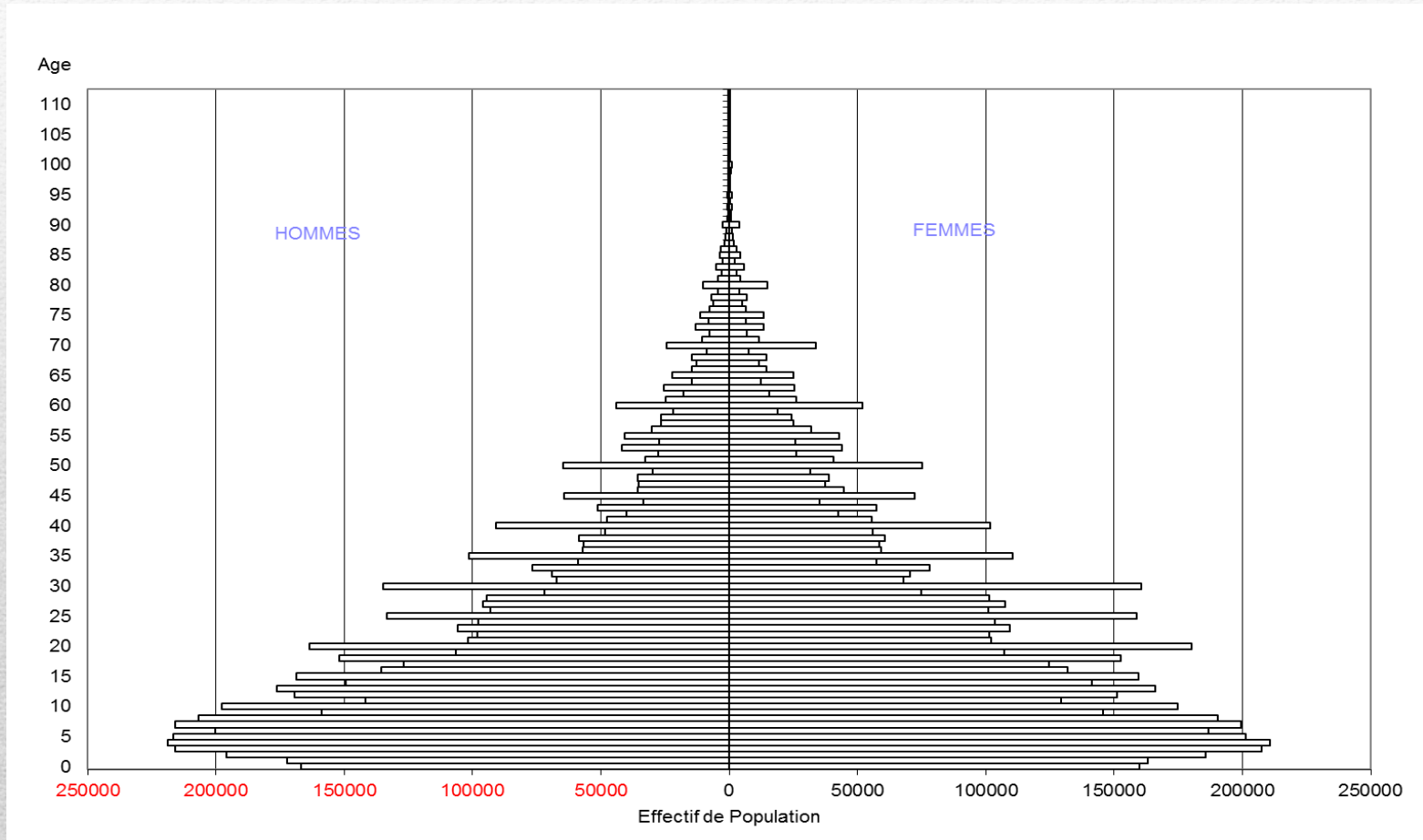
- ❑ Sous/sur dénombrement
- ❑ Erreur de déclaration des âges (yc préférence pour 0 ou 5)
- ❑ Changements dans le temps de la fécondité, mortalité, migration
  - si fécondité baisse, moins d'enfants aux jeunes âges

Donc, des investigations supplémentaires sont nécessaires...

---

# Analyse des distributions par sexe et âge

## Pyramide des âges avec préférences 0 et 5



# Analyse des distributions par sexe et âge

## Contrôles de base

- Examiner les procédures
    - D'évaluation de la qualité (y compris enq post-censitaire)
    - De corrections (manuels et algorithmes de corrections)
      - Imputation des données manquantes : % / variable ?
  - Vérifier les procédures de saisie
    - Saisie manuelle (erreurs de saisie)
    - Lecture optique (confusion 1 et 7, pb questionnaires sales)
    - Saisie à l'enquête (complétude, supervision limitée)
-

# Analyse des distributions par sexe et âge

## Autres contrôles simples

- Comparer les résultats de recensement avec toute autres données disponibles
    - Sources non démographiques se rapportant à celle-ci
  - Comparer la distribution de la population avec des caractéristiques connues :
    - Densité de population rurale/urbaine
  - Comparer les effectifs de population et de ménages
    - Taille moyenne des ménages
    - Proportions de ménages isolés
    - Par région et rural/urbain
  - Porter une attention toute particulière aux groupes jugés difficile à dénombrer
    - Très jeunes enfants, zones d'accès difficile, ...
-

# Analyse des distributions par sexe et âge

- Cohérence externe

- Comparaison à des données antérieures (évolutions cohérentes, taux d'accroissement selon les 2 méthodes – sur la population totale ou sur certains groupes)
  - Comparaison avec une population similaire
  - Comparaison avec des données de projections ou d'état civil
  - Comparaison avec des sources administratives
  - Comparaison avec des données d'observatoires de population
  - etc.
-



# Exemples

---

# Analyse des distributions par sexe et âge

Basée sur 1 seule source

1. Identification des âges attractifs
    - distribution par sexe et âge
  2. Différences de couverture ou d'omission
    - rapports d'âge
    - rapport de masculinité
-



# Analyse des distributions par sexe et âge

Basée sur 1 seule source

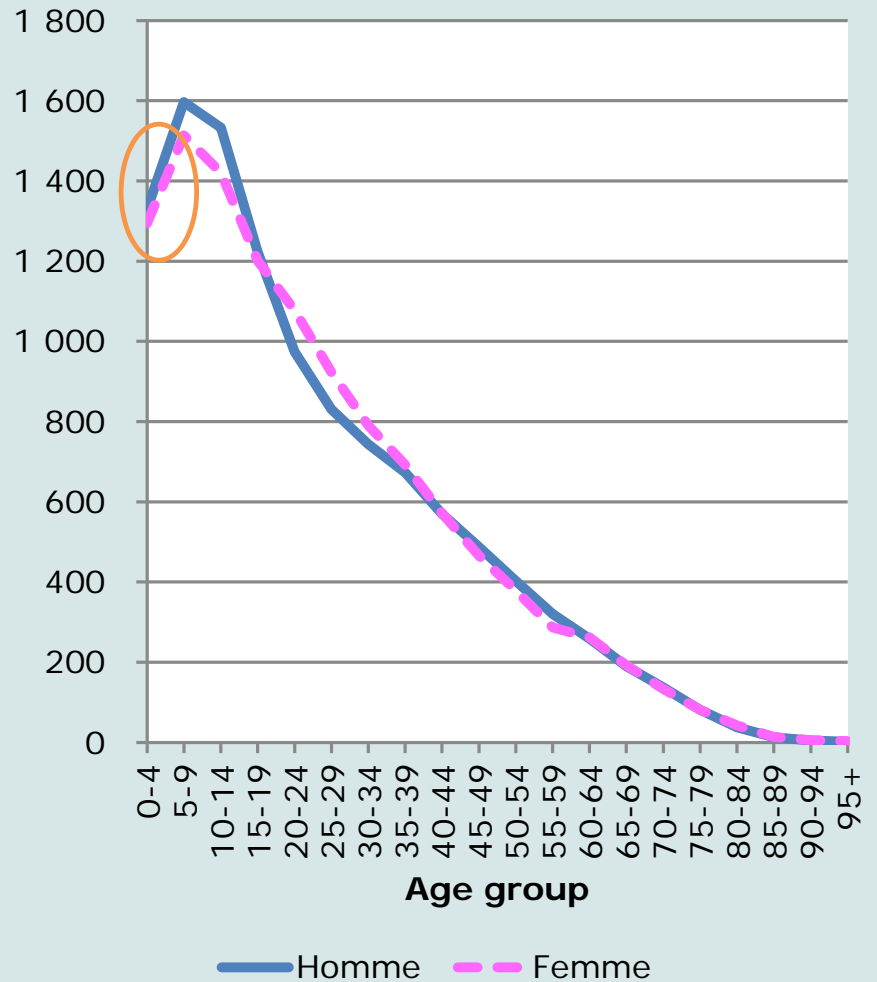
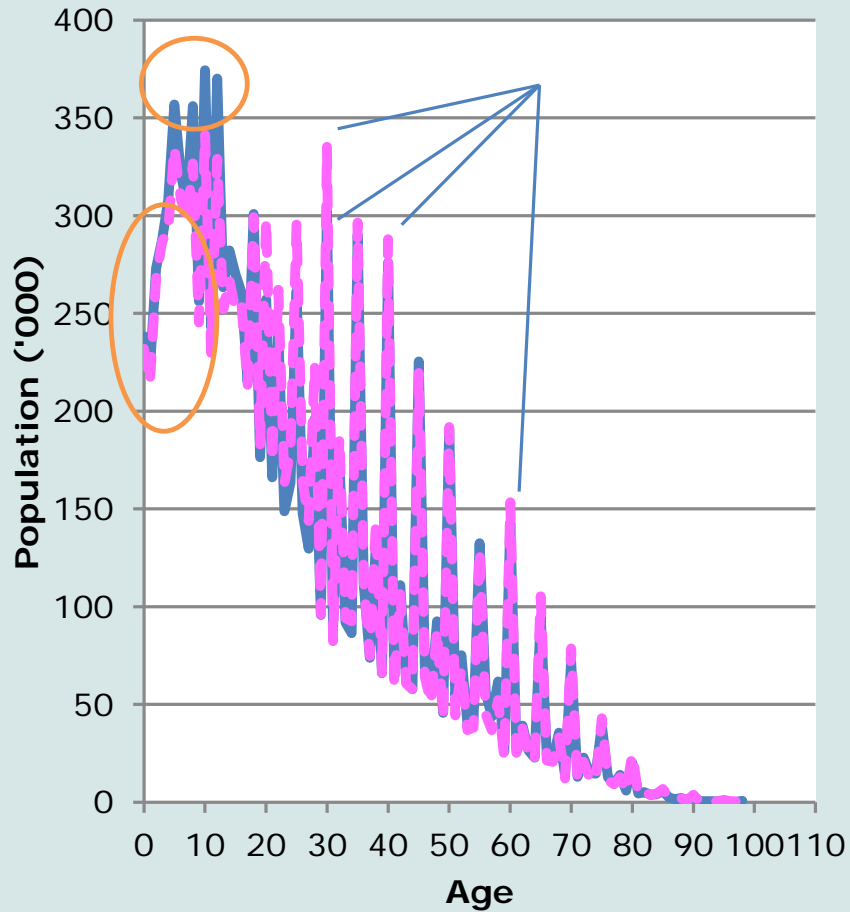
- Par années d'âge et par groupes d'âges :
  - Graphique par âge continu plutôt que pyramide
  - Constat visuel des attractions
  - Indices de Myers, indice de Whipple, indice de régularité sexe-âge des NU

*« Ces procédures sont utiles comme mesures synthétiques ou pour des fins comparatives, mais ne donnent généralement pas d'idée sur les schémas d'erreur dans les données qui ne puisse être tirées des méthodes graphiques ou du calcul des ratios » (US Bureau of Census, 1985)*

➔ On s'attend à une évolution régulière

---

# Age-sex distribution of the population of Nepal, 2001 census



# Analyse des distributions par sexe et âge

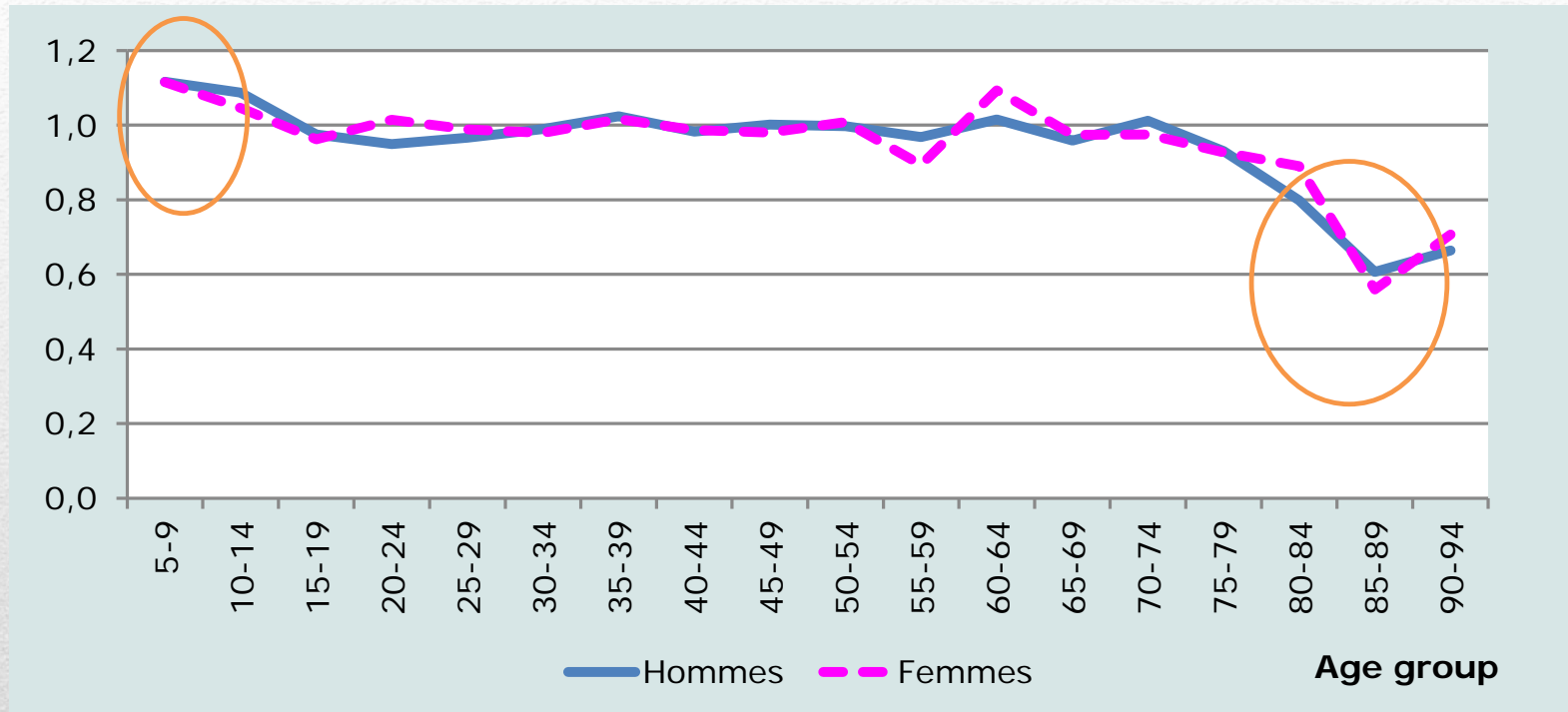
## Rapport d'âge

*Le rapport d'âge pour un groupe donné est le ratio du double de la population dans ce groupe d'âge à la somme de la population dans chacun des groupes d'âge adjacents (encadrants).*

$${}_nAR_x = \frac{2 \cdot {}_nN_x}{({}_nN_{x-5} + {}_nN_{x+5})} * 100$$

- Sous-dénombrement
  - Déplacement entre groupes d'âge
    - Si variation linéaire entre groupes d'âge,  ${}_nAR_x$  proche de 1 (sauf facteurs extérieurs : migration, guerre, ...)
    - Entre naissances et âges adultes (vers 45 ans),  ${}_nAR_x$  recule lentement
    - Après 45 ans, baisse rapide de  ${}_nAR_x$  du fait d'une surmortalité masculine
-

# Rapports d'âge et groupes d'âge, Nepal recensement 2001



- $nSR_x$  proche de l'unité pour tous sexes, sauf aux âges jeunes (omissions chez 0-4 ans, et déplacement dans 5-9 ans)
- Aux âges avancés, chute des  $nSR_x$  plus conformes aux attentes (mortalité)
- Effectifs à comparer avec données de recensement passé, estimations et/ou état civil, etc.

# Analyse des distributions par sexe et âge

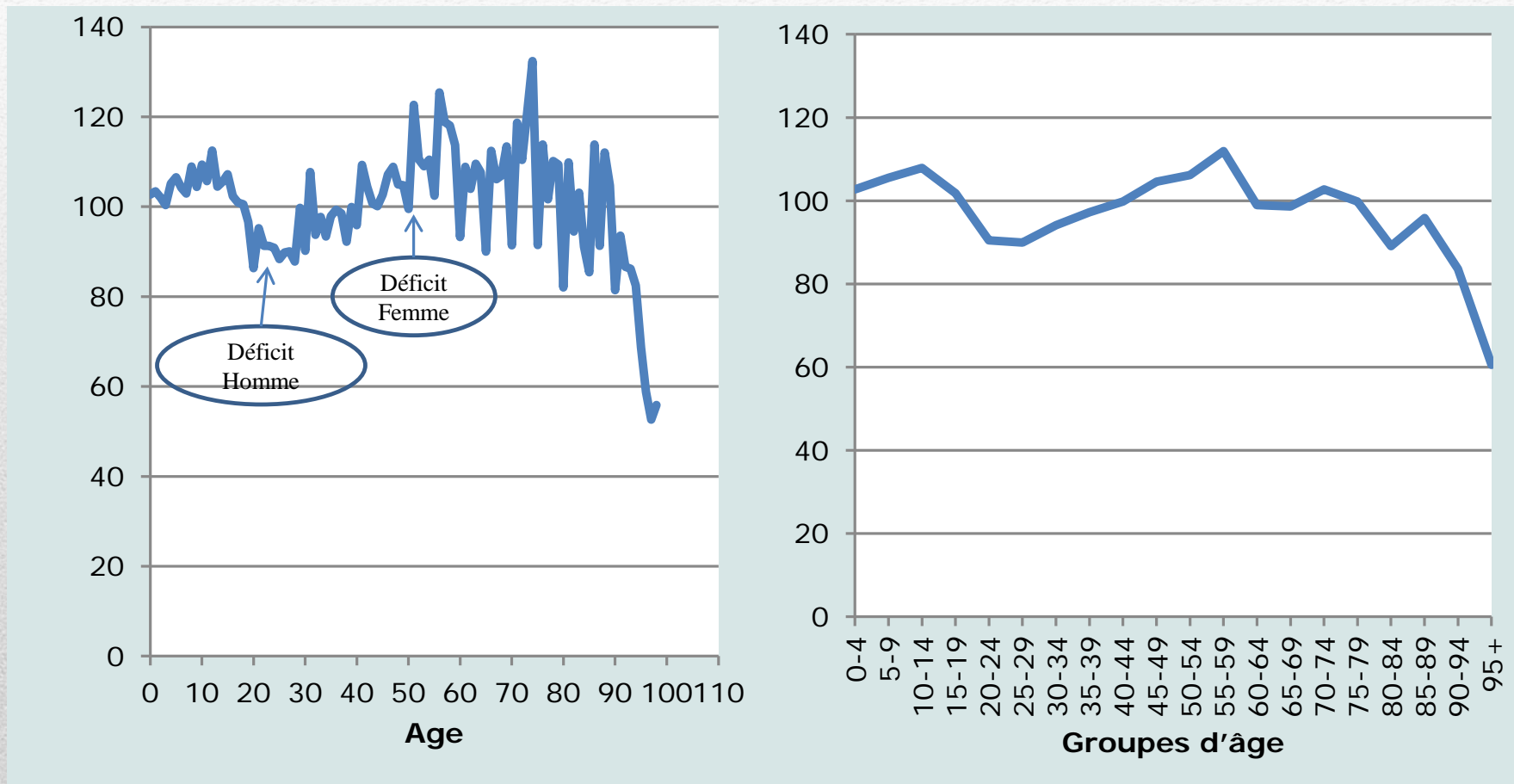
## Rapports de masculinité

- Le rapport de masculinité global est le rapport du nombre d'hommes pour 100 femmes dans la population
- Il peut être calculé sur les naissances et dans un groupe d'âge particulier

$${}_nSR_x = \frac{{}_nN_x^m}{{}_nN_x^f} * 100$$

- on s'attend à un rapport de 100 à 105 à la naissance
  - le SR doit refléter les différences de mortalité selon le sexe (âge à la maternité, âge élevé)
  - si pas de différence de mortalité  $\sim < 100$
  - d'autant plus faible que le population est âgée
-

# Rapports de masculinité par année et groupes d'âge, Nepal recensement 2001



→ Résultats à rapprocher à des données récentes (rapports d'âge et sexe)

# Analyse des distributions par sexe et âge

Test par comparaison avec une autre source

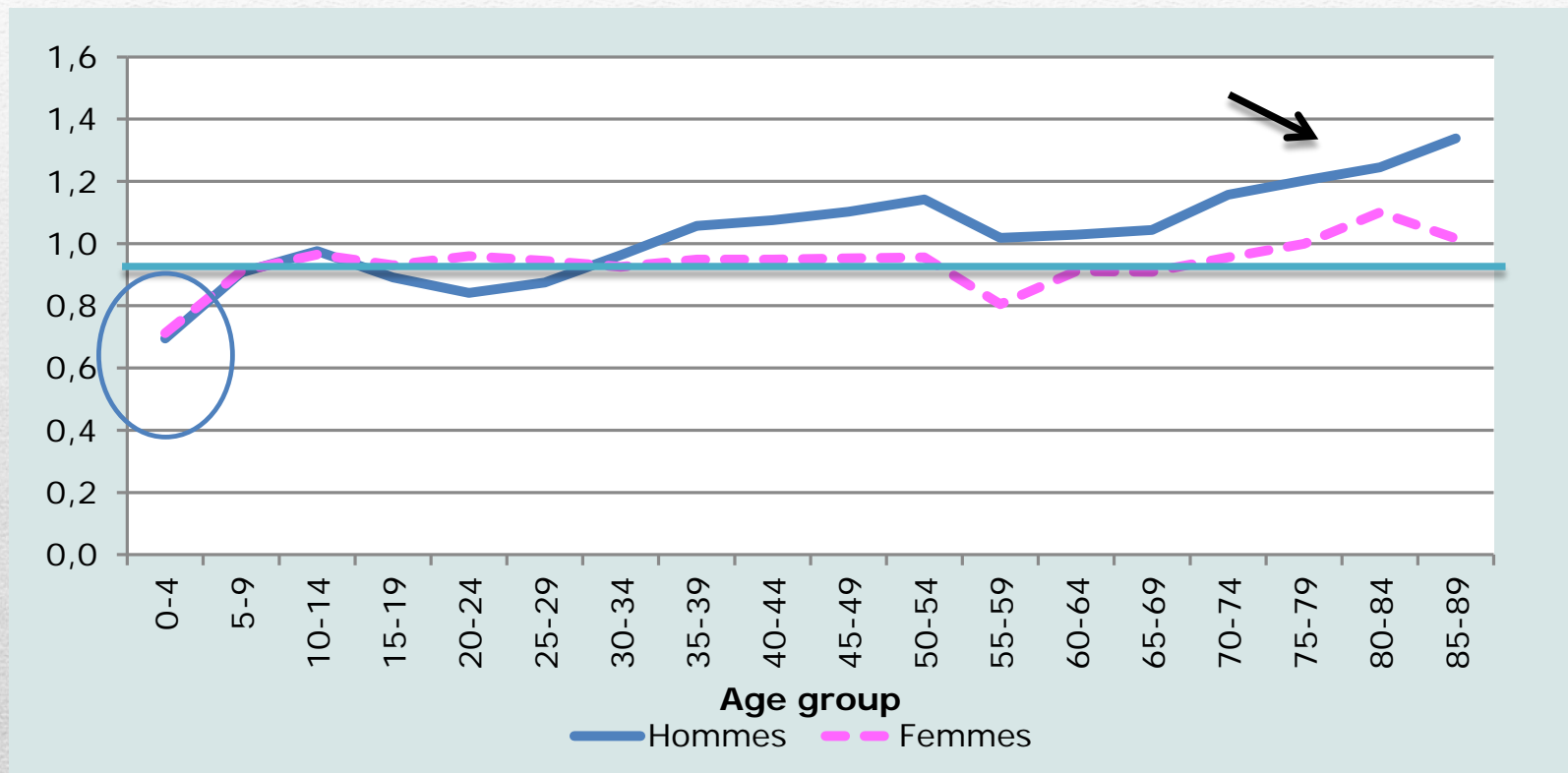
- Comparaison aux estimations les plus récentes

$${}_nR_x = \frac{{}_nN_x^m S1}{{}_nN_x^m S2} * 100$$

$${}_nR_x = \frac{{}_nN_x^f S1}{{}_nN_x^f S2} * 100$$

---

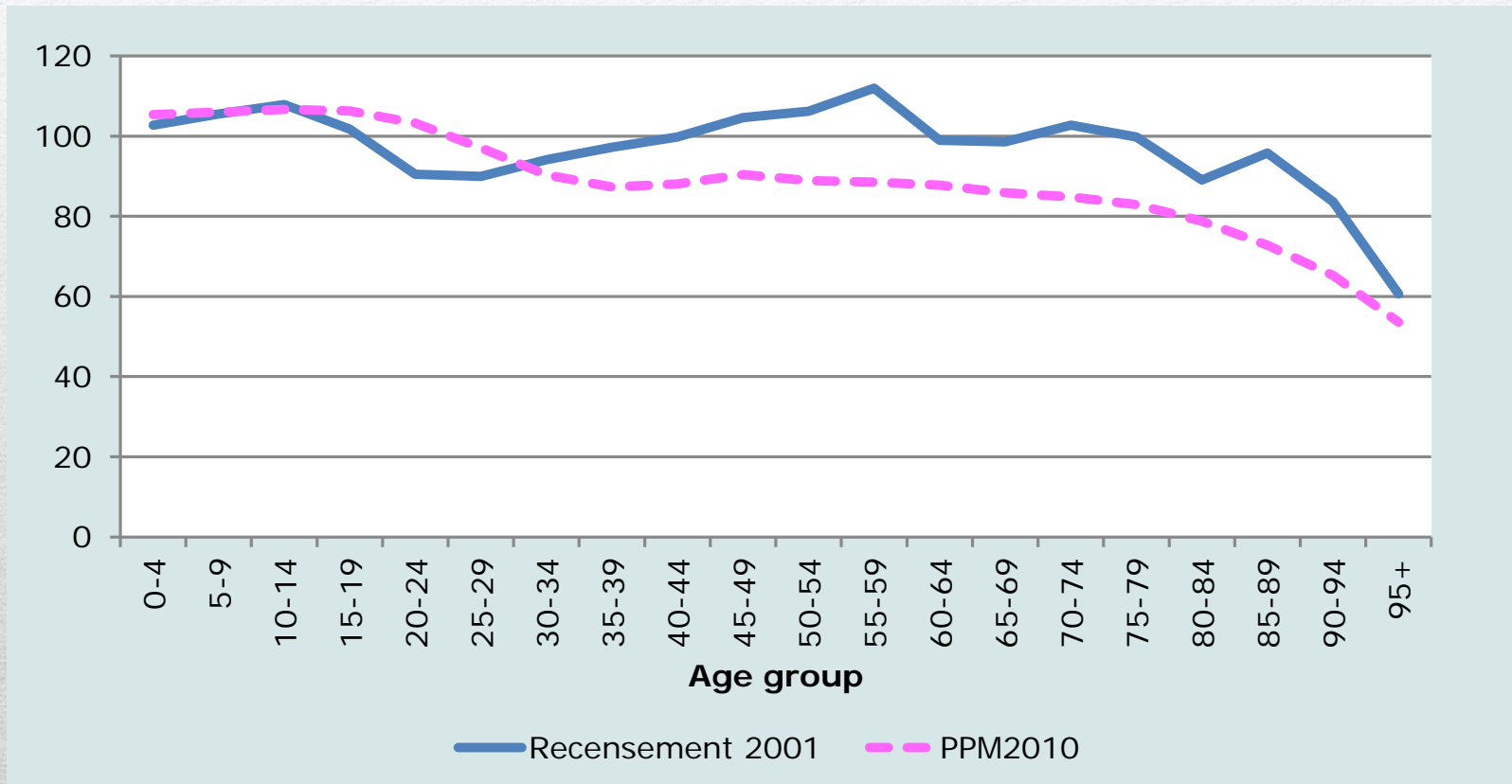
# Rapport par âge et sexe de la population dénombrée en 2001 à la population estimée par la Division de la population des Nations Unies pour 2001, Népal



Par rapport PPNU, population des hommes surestimée



# Rapports de masculinité par groupe d'âge dans la population dénombrée en 2001 et la population estimée par la Divion de la population des Nations Unies pour 2001, Népal



➤ Différences importantes entre les  ${}_nR_x$  par âge RGPH2001 vs PPM2010

# Analyse des distributions par sexe et âge

## Basée sur plusieurs recensements

- Principes de base :
    - les variations de population doivent être régulières
    - les écarts importants doivent pouvoir s'expliquer par des événements connus
    - Les écarts non expliqués sont signes d'erreurs potentielles
  - Méthodes fondées sur l'estimation de la mortalité adulte :
    - Taux de croissance intercensitaires
    - Rapports de survie de cohortes
-

# Analyse des distributions par sexe et âge

## Basée sur plusieurs recensements

### Taux de croissance intercensitaires

$$r = \frac{\ln\left(\frac{N(t_2)}{N(t_1)}\right)}{(t_2 - t_1)}$$

- $N(t_1)$  et  $N(t_2)$  représentent populations totales aux temps respectivement  $t_1$  et  $t_2$
  - sans mobilité (croissance naturelle),  $0 < r < 3,5\%$
  - Une forte croissance de  $r$  traduit peut indiquer un sous dénombrement des recensements successifs
-

# Analyse des distributions par sexe et âge

## Basée sur plusieurs recensements

### Rapports de survie des cohortes (CSR)

$${}_nCSR_x(a) = \frac{{}_nN_{x+a}(t+a)}{{}_nN_x(t)}$$

### Ratio de survie (nRx)

$${}_nR_x = \frac{{}_nN_{x+a}(t+a)/{}_nN_x(t)}{{}_nL_{x+a}/{}_nL_x}$$

- Cohorte = groupe d'âge  
→ absence de migration  $Nt_2 < Nt_1$
  - Taux/Rapport de survie (CSR) :  $CSR_f > CSR_h$  et  $CSR < 1$
  - Conditions : absence de migrations importantes, frontières immuables, même population recensée
-

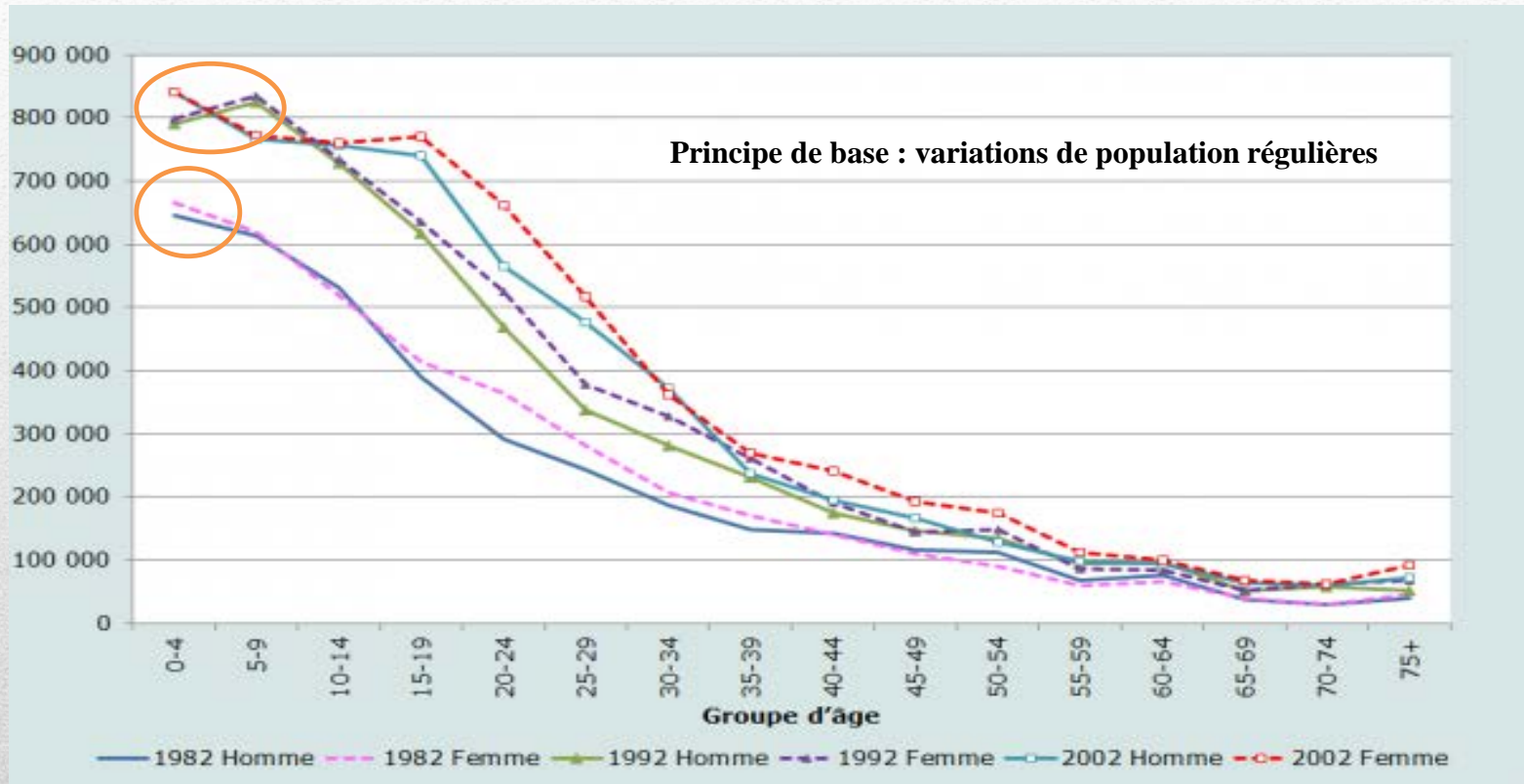
# Tableau 1 Population du Zimbabwe par âge et sexe, recensements de 1982, 1992 et 2002

Age	1982		1992		2002	
	Homme	Femme	Homme	Femme	Homme	Femme
0	133 070	136 960	167 552	169 064	170 054	170 277
1-4	510 260	528 390	621 411	626 664	668 008	667 730
5-9	612 760	619 300	821 319	832 469	764 453	769 247
10-14	529 750	518 740	724 905	731 846	754 587	757 657
15-19	390 160	412 610	615 728	632 510	736 686	766 890
20-24	290 380	364 200	466 837	523 060	564 034	658 873
25-29	243 420	281 060	335 713	376 495	473 984	513 793
30-34	185 400	206 760	280 066	326 299	369 836	360 291
35-39	147 920	170 170	229 360	259 555	235 692	268,797
40-44	142 050	139 530	174 266	189 509	194 702	239 727
45-49	116 490	110 390	145 437	143 441	165 437	191 168
50-54	111 780	90 880	133 261	147 339	128 029	173 229
55-59	67 400	60 800	94 713	86 729	98 417	112 498
60-64	76 850	65 260	95 510	84 213	94 447	99 420
65-69	38 810	38 860	51 202	50 902	64 301	67 851
70-74	29 810	30 500	58 279	62 479	60 311	62 464
75+	39 410	46 760	52 026	68 403	71 950	92 311
Inconnu	7 900	6 680	15 952	18 034	19 252	25 254
Total	3 673 620	3 827 850	5 083 537	5 329 011	5 634 180	5 997 477

## Tableau 2 Population ajustée du Zimbabwe par âge et sexe, recensements de 1982, 1992 et 2002

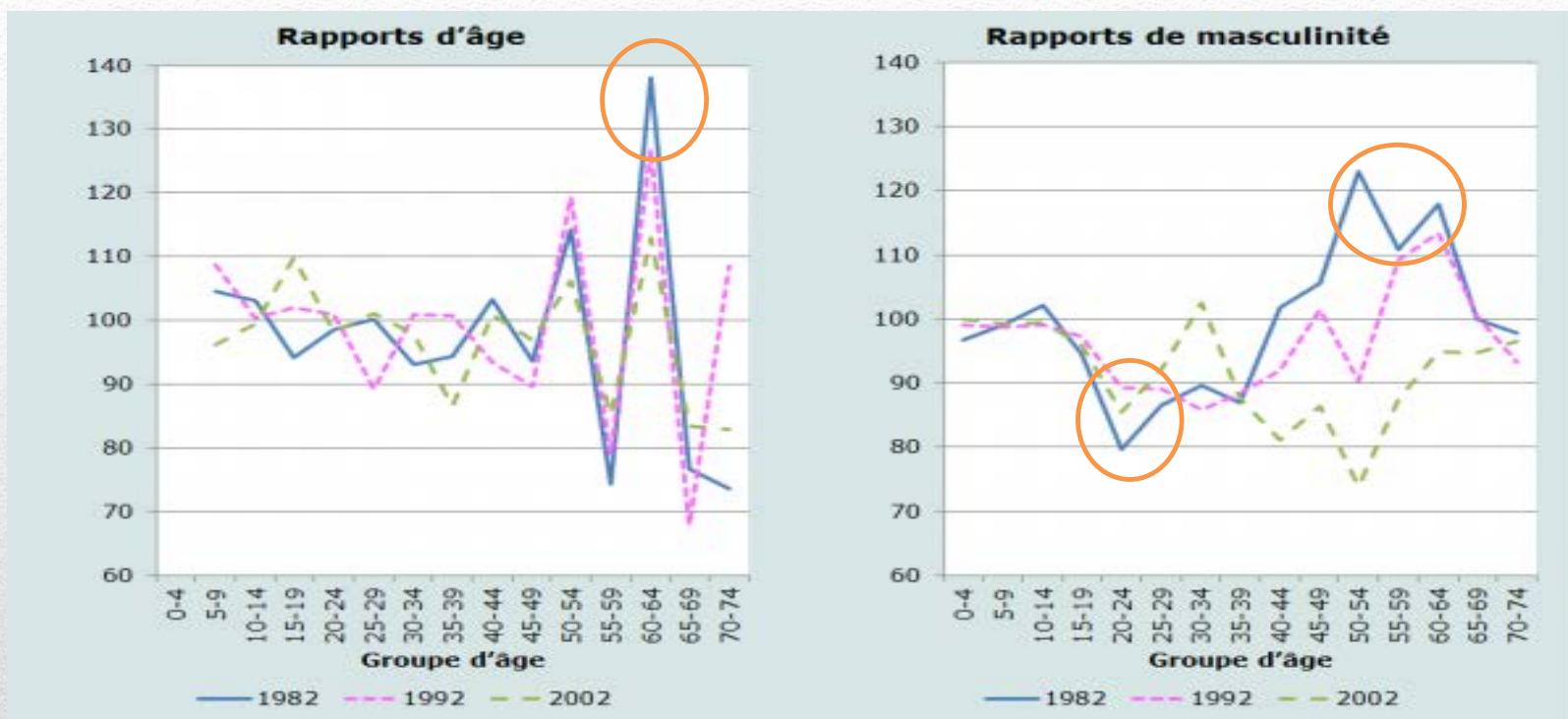
Age	1982		1992		2002	
	Homme	Femme	Homme	Femme	Homme	Femme
0	133 357	137 199	168 079	169 638	170 637	170 997
1-4	511 360	529 314	623 367	628 792	670 298	670 554
5-9	614 081	620 383	823 904	835 296	767 074	772 500
10-14	530 892	519 647	727 187	734 331	757 174	760 861
15-19	391 001	413 331	617 666	634 658	739 212	770 133
20-24	291 006	364 837	468 307	524 836	565 968	661 659
25-29	243 945	281 551	336 770	377 773	475 609	515 966
30-34	185 800	207 121	280 948	327 407	371 104	361 815
35-39	148 239	170 467	230 082	260 436	236 500	269 934
40-44	142 356	139 774	174 815	190 152	195 370	240 741
45-49	116 741	110 583	145 895	143 928	166 004	191 976
50-54	112 021	91 039	133 680	147 839	128 468	173 962
55-59	67 545	60 906	95 011	87 023	98 754	112 974
60-64	77 016	65 374	95 811	84 499	94 771	99 840
65-69	38 894	38 928	51 363	51 075	64 521	68 138
70-74	29 874	30 553	58 462	62 691	60 518	62 728
75+	39 495	46 842	52 190	68 635	72 197	92 701
Inconnu						
Total	3 673 620	3 827 850	5 083 537	5 329 011	5 634 180	5 997 477

## Distribution de la population du Zimbabwe par âge et sexe, recensement de 1982, 1992 et 2002



- Pour les 3 Rgph, excédant des femmes entre 15-35 ans (migration de travail ?)
- Sous dénombrement des moins de 5 ans en 1992 (comparés au 5-9 ans)

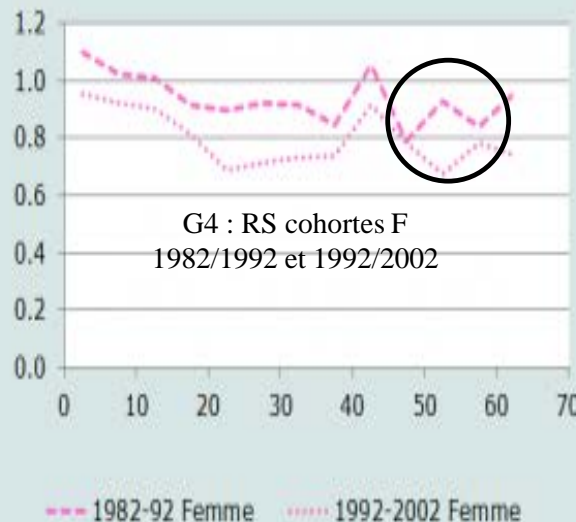
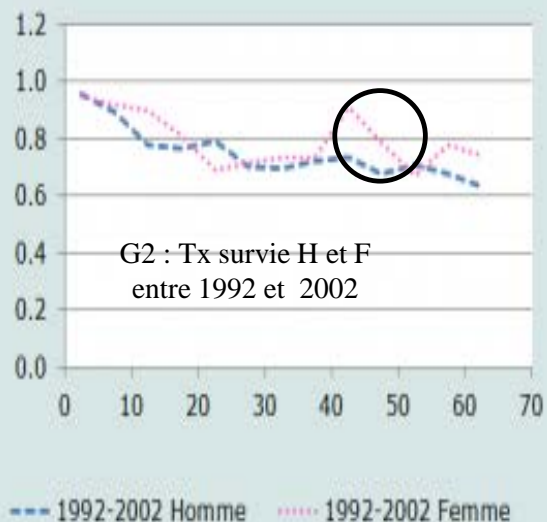
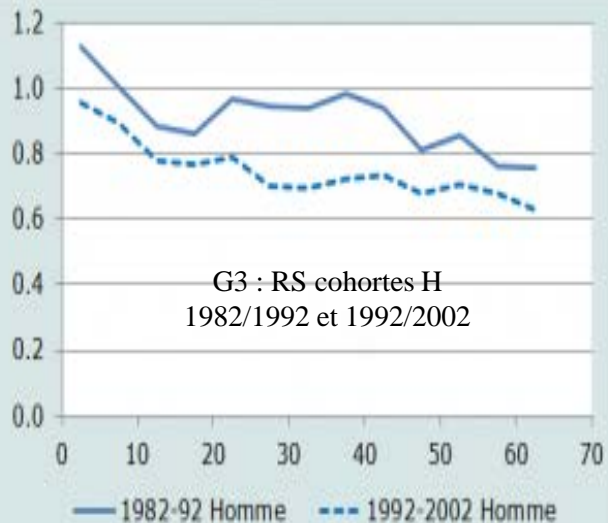
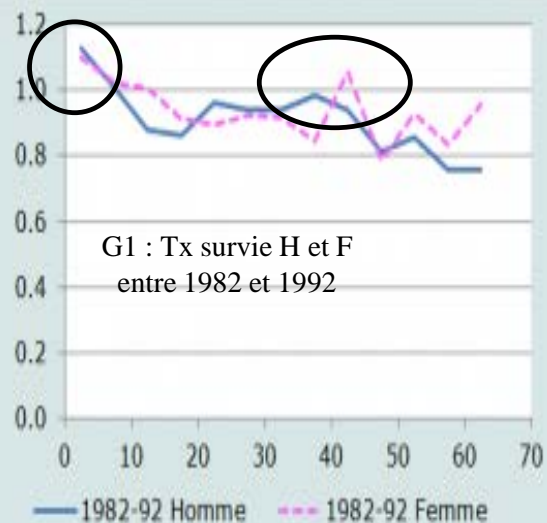
# Rapports d'âge et rapports de masculinité, Zimbabwe



- Rapports d'âge du groupe 60-64 ans très élevés dans les 3 Rgph
- Impact sur les groupes 55-59 et 65-69 ans
- RM proches de 100 avant 15 ans, et baisse au-delà (migrations des jeunes hommes)
- Hausse douteuse du RM entre 35-55 ans en 1992 (sous déclaration des femmes ?)
- RM plus en faveur des hommes, signe d'un vieillissement systématique de leur âge



# Rapports de survie des cohortes par âge et sexe, recensements du Zimbabwe 1992 et 2002



## Rapport de survie

- Décroissance du taux au fil des âges / RSh vs RSf
- Hausse de la mortalité au fil des âges (effet vih)
- G1 et 2 : Sous dénombrement des enfants des 2 sexes jusqu'à 20 ans en 1982 (Forte immigration ?)
- Décroissance des RS (forte mortalité) à tous âges

## Rapport de survie des cohortes

- Importante hausse de la mortalité (effet vih)
- RSC erratique, qualité données ?
- Accroissement mortalité des enfants dû au sous dénombrement de 1982

# Tests à partir des données d'observatoires de population

- Principes de base :
    - S'assurer de l'harmonisation de la définition des concepts (notamment la définition du ménage)
    - Extraire les données aux bonnes dates (cf. dates de recensement)
    - Les districts de recensements couvrant les zones à comparer doivent être bien identifiées et délimitées.
    - Traitements des visiteurs, ménages collectifs et migrants saisonniers (à exclure des HDSS)
  - Méthodes:
    - Approche enquête post censitaire
    - Rapports de survie de cohortes
-

# Tests à partir de données d'observatoires de population

## Population supplémentaire dénombrée

### Niakhar :

- les travailleurs saisonniers Norane (y compris les bonnes) et Navétane même s'ils ont passé plus de 6 mois hors de la zone;
- les travailleurs ayant une épouse ou des enfants dans la zone même s'ils ont passé plus de 6 mois hors de la zone ;
- les élèves ou talibés sortis de la zone pour raison d'études.

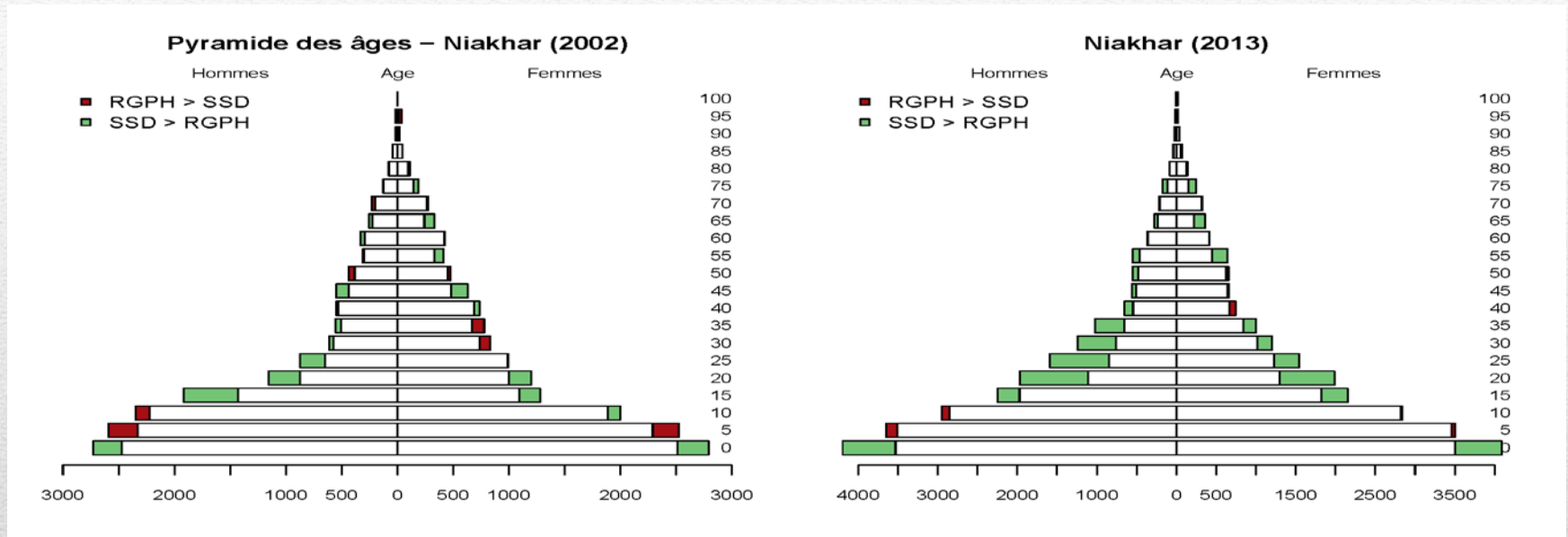
### Bandafassi

- Personnes absentes depuis moins de quatre ans

### Mlomp

- Personnes absentes depuis moins de deux ans
-

# Structure par âge de la population Niakhar selon RGPH et HDSS en 2002 et 2013



➤ Relatif bon accord entre sources, sauf chez les hommes adultes (20-39). Sans doute à nouveau l'effet des règles de résidence.

- Les migrants inclus dans la population DSS en 2013 (et moins en 2002).
- Problèmes avec les 5-9 ans et 10-14.

# Structure par âge de la population Niakhar selon RGPH et HDSS en 2002 et 2013

## Qualité des déclarations d'âge (niveau agrégé) :

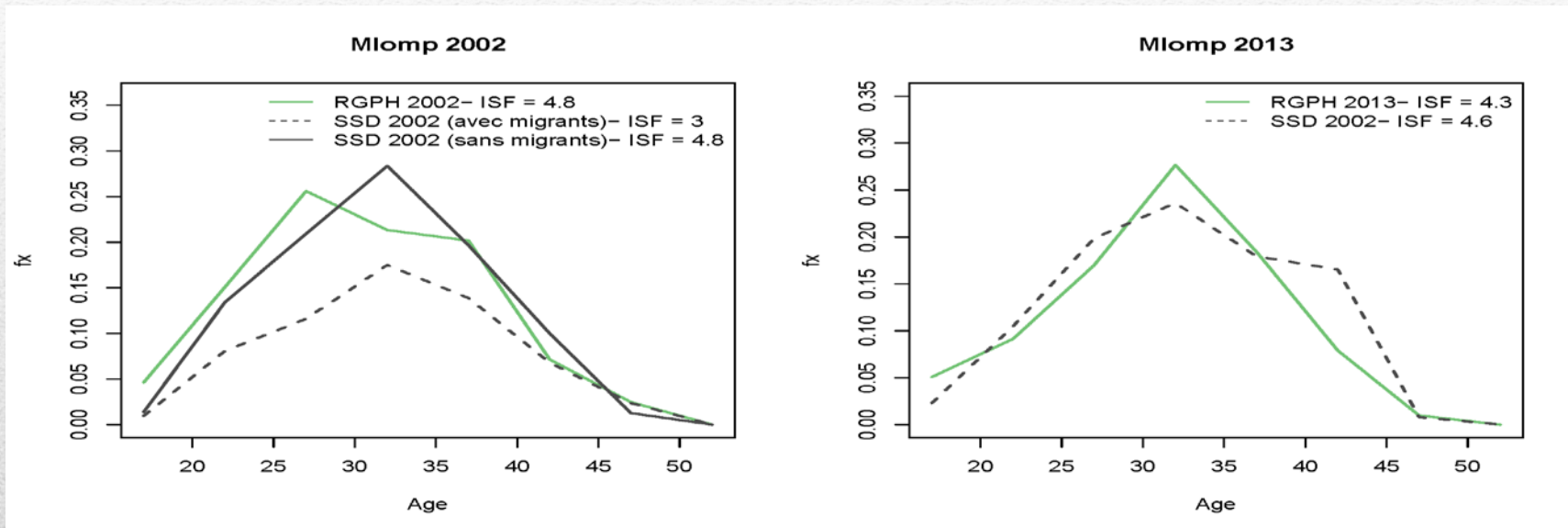
**Indice de Myers** (permet de déceler les préférences ou les répulsions des âges se terminant par les chiffres allant de 0 à 9, soit une estimation de la proportion minimale de personnes pour qui l'âge est mal enregistré):

	2002			2013		
	Mlomp	Bandafassi	Niakhar	Mlomp	Bandafassi	Niakhar
HDSS	2.6	2.7	2.8	2.4	1.9	1.9
RGPH	7.7	16.1	19.0	5.3	6.5	8.8

Déclaration des âges de qualité proche dans les HDSS, et surtout nette amélioration entre 2002 et 2013 aussi bien dans les HDSS que dans les recensements.

# Comparaisons indicateurs de fécondité selon RGPH et HDSS en 2002 et 2013

Taux de fécondité estimés en 2002 et 2013 à Mlomp



- Migrants à pour le calcul des TFA, sinon comparaison biaisée.
- Les femmes enregistrées comme résidentes au recensement sont nettement plus fécondes que les autres.

Quand les migrantes sont exclues du calcul, l'observatoire de Mlomp procure des taux de fécondité très proches de ceux obtenus à partir des recensements.

# Comparaisons indicateurs de mortalité selon RGPH et HDSS en 2002 et 2013

## Mortalité des enfants

- Tendances de la mortalité obtenues à partir du nombre d'enfants nés vivants et survivants des femmes adultes en 2013 (méthode classique de Trussell, Hill, 2013).
- Structure par âge (modèle Sud de tables Coale-Demeny pour HDSS, Masquelier et al., 2014).

## Mortalité des adultes

- Probabilités de décès obtenus en appliquant la méthode des orphelins (Brass and Bangboye, 1981; Timæus, 1992).
- Modèle Sud de Coale-Demeny utilisé pour convertir les différentes probabilités de survies ( $np_{25}$  chez les femmes et  $np_{35}$  chez les hommes) en un indice unique de mortalité entre 15 et 60 ans ( $45p_{15}$ ).
- Estimations indirectes lissées à l'aide d'une moyenne mobile sur 2 ans.

## Taux de mortalité par âge

- Compte tenu des faibles effectifs, pas de méthodes de distribution des décès (Méthode générale de l'accroissement démographique et méthode de l'extinction des cohortes synthétiques (Hill, 1987; Bennett and Horiuchi, 1984).
  - Calcul direct avec décès déclarés dans les 12 mois ayant précédé le Rgph (indicateurs standards  $1q_0$ ,  $5q_0$ ,  $45q_{15}$ ,  $e_0$  et  $e_5$ ).
  - Pas de valeurs présentées pour Mlomp en 2013 (pb d'effectifs), mais calcul possible en 2002 (naufrage du Joola).
-

## Comparaisons indicateurs de mortalité selon RGPH et HDSS en 2002 et 2013

Année		Bandafassi		Mlomp		Niakhar	
		HDSS	RGPH	HDSS	RGPH	HDSS	RGPH
2002	1q0	0.11	0.04	0.05	>0.01	0.06	0.02
	4q1	0.09	0.11	0.04	0.05	0.08	0.08
	45q15	0.25	0.21	0.54	0.59	0.26	0.24
	20q60	0.68	0.71	0.64	0.43	0.52	0.49
	e <sub>0</sub>	54.3	58.4	46.6	46.3	59.6	62.8
	e <sub>5</sub>	61.9	62.9	45.7	43.5	64	64.5
2013	1q0	0.06	0.08			0.02	0.02
	4q1	0.04	0.04			0.02	0.03
	45q15	0.29	0.38			0.13	0.19
	20q60	0.59	0.48			0.52	0.52
	e <sub>0</sub>	61.1	58.8			70.7	68.7
	e <sub>5</sub>	62.4	61.3			68.7	66.6



# Comparaisons indicateurs de mortalité et de fécondité selon RGPH et HDSS en 2002 et 2013

## Tables de mortalité

- Sous-estimation sensible de la mortalité infantile dans le recensement de 2002, mais pas en 2013.
  - La mortalité juvénile est partout assez bien estimée (peut-être sous l'effet d'une attirance sur 1 an)
  - Entre 15 et 60 ans les données du recensement tendent à procurer des niveaux de mortalité très proches de ceux enregistrés dans le suivi, voire même légèrement plus hauts en 2013.
  - Entre 60 et 80 ans, compte tenu des omissions plus fréquentes des personnes âgées (et de leurs décès) ainsi que des erreurs sur les âges, la concordance entre les estimations du recensement et celles du suivi est également remarquable.
-

# Enquêtes post-censitaires

- ❑ Retour dans des zones-échantillons pour ré-administrer un questionnaire court
  - ❑ Appariement des individus et des ménages
  - ❑ 2 hypothèses
    - Indépendance des probabilités d'échapper au recensement et d'échapper à l'EPC (souvent fausse)
    - Identification des mêmes individus dans les deux enquêtes (fausse si population mobile ou intervalle long)
- ➔ Prendre connaissance du rapport de l'EPC pour en comprendre les éventuelles faiblesses
-

MERCI DE VOTRE ATTENTION

---