Mission to Kigali, Rwanda
September 9 to 21, 2004

Juan Muñoz[1], consultant for Oxford Policy Management
October 2004

October 2004

---

[1].Address: Sistemas Integrales - Casilla 13168 – Santiago, Chile. Tel: (56-2) 638-1841 and 639-4554. Fax: (56-2) 639-2086. E-mail: juan.munoz@ariel.cl

By appointment with Oxford Policy Management Ltd (OPM,) I visited Kigali from September 9 to 21, 2004, to help the Rwandan *Direction de la Statistique* (DS) in various activities related to the design and implementation of the Integrated Living Conditions Survey (*Enquête Intégrale sur les Conditions de Vie*, EICV2,) to be fielded in 2005, the Baseline Rural Sector Survey (BRSS) and a series of Light Rural Sector Surveys (LRSSs,) to be fielded from October 2004 onwards. My terms of reference requested me to

- peer review the sample designs proposed by the sampling expert (for OPM), and make recommendations on possible amendments to the DS and OPM team;

- work with the teams on data management systems for the surveys; and

- participate in a workshop for stakeholders to review the EICV1 and make recommendations for the conduct of the EICV2.

I met with Mr Philip Gafishi, Director of Statistics, with senior technical staff of the Direction de la Statistique, and with Mr Paul-Henri Wirrankoski, OPM's resident advisor to the DS. My mission was concurrent with the visits to Kigali of OPM experts Mary Strode, Cynthia Donovan, Andrew McKay and Timothy Jones, with whom I worked in close collaboration.

In addition to the tasks specified in my terms of reference, I revised with Mr Gafishi the schedule of preparatory activities and the budget for the EICV2.

The rest of this report focuses on sampling, with the intention of rendering it a self-standing document on this topic. Appendixes 1 and 2 present the draft schedule and budget prepared with Mr Gafishi and his team, and summarize my presentations in the stakeholders' workshop.

Sampling

Peer review of the EICV2 and rural sector surveys sample designs

The sample designs for the EICV2 and for the program of rural sector surveys proposed by OPM's expert David Megill last June can be briefly summarized as follows:

- The sample proposed for the EICV2[2] is very similar to the one used by the EICV1[3]. It allocates a sample of 6,900 households to thirteen explicit strata: 900 households to Kigali Town, 720 households to the other urban areas, and 480 households to the rural parts of each of Rwanda's other eleven provinces. Within each stratum the sample is selected in two stages: In the first stage, using the 2002 Census *Zones de Dénombrement* (ZDs) as Primary Sampling Units (PSUs) and the total number of households as a measure of size, a certain number of PSUs were selected with probability proportional to size (*pps*,) with implicit

---

[2] Megill, D. *Recommendations on sample design and estimation methodology for the Rwanda Enquête Intégrale sur les Conditions de Vie des Ménages 2005*. Kigali, June 2004.

[3] Scott, C. *Enquête Intégrale sur les Conditions de Vie des Ménages 2000 (avec volet Budget-Consommation) – Plan de Sondage*. Kigali, July 2000.

stratification by urban/rural and by an indicator of wellbeing in urban areas. In the second stage, after a household listing operation to be conducted in all chosen PSUs, a fixed number of households (nine in Kigali and the other urban areas; twelve in the rural areas) are to be selected with systematic equal-probability (*sep*) sampling in each PSU.

- The design for the program of rural sector surveys[4] defines eleven explicit strata, Kigali ngali and Kigali Town (considered as a single stratum,) and each of Rwanda's other ten provinces. A total sample of 5,760 agricultural households (referred to as *farmers* from here onwards for simplicity) is allocated among these strata approximately in proportion to the number of farmers, with sample sizes ranging from 384 to 624 farmers per stratum. Within each stratum, the sample is selected in two stages. In the first stage, using Agricultural ZDs[5] as PSUs and the number of farmers as a measure of size, a certain number of PSUs is selected with *pps*. In the second stage, after a household listing operation to be conducted in all chosen PSUs, twelve farmers are to be selected with *sep* in each PSU. Other than the exclusion of non-farmers, no stratification is considered at the second stage. The full 5,760-farmer sample will be used by the *Baseline Rural Sector Survey* (BRSS,) to be fielded in 2006. This sample is divided into four random sub-samples of 1,440 farmers each, one of which was chosen as the panel for the series of annual *Light Rural Sector Surveys* (LRSSs,) to be fielded from October 2004 onwards.

The two sample designs are mutually independent, and independent of other samples previously conducted in Rwanda, including the EICV1. In other words, the EICV1 and the EICV2 are conceived as two independent cross-sectional samples, not as panel.

Had the conditions prevailing at the time of Mr Megill's visit remained unchanged, I would be unable to make any significant comments or additions to his proposal. The two designs are simple and solid, and they would serve most of the purposes of their respective surveys well. The only apparent shortcoming of the rural sector sample is that the sampling errors in the estimation of the production of crops harvested by a minority of the farmers (such as Irish potatoes,) may be too large. This problem could be approached, in principle, either by increasing the overall sample size or by implementing some form of stratification at the second sampling stage (in order to select these farmers preferentially.) However, as Mr Megill correctly points out, both options present practical problems. The total sample size would have to be much larger, not just marginally larger, in order to have a significant effect on sampling errors, and it could well be counter-productive as a result of the larger non-sampling errors inherent to such a large operation, even in the unlikely case it could be afforded. The case against second-stage stratification is based on subtler grounds: with second-stage stratification the household listing operation becomes a mini-survey rather than a simple enumeration of all households in each PSU; this often shifts the interest of the field staff from listing to interviewing and may result in serious selection biases, since the households that are (or seem to be) "harder to interview" may end up being excluded. An additional difficulty of

---

[4] Megill, D. *Recommendations on sample design and estimation methodology for Rwanda Agricultural Surveys 2004*. Kigali, June 2004.

[5] "Agricultural ZDs" were operationally defined as those with at least 70 percent of farmers. This obviously includes most of the ZDs formally qualified as "rural", but also many of those qualified as "urban".

second-stage stratification is that it renders the selection of the farmers to be interviewed – which is relatively straightforward in the absence of stratification – a much more complicated operation that almost always results in errors when it is entrusted to the enumerators themselves.

In brief, the designs proposed by Mr Megill in June 2004 seem adequate to me. If the resulting sampling errors for the production of Irish potatoes are judged to be unacceptably large, I would resort to second-stage stratification, but with two important precautions: [a] The household listing operation needs to be very well supervised, to minimize the risk of excluding any farmers, and [b] The random selection of the farmers to be interviewed should not be entrusted to the enumerators themselves; the listing forms should instead be brought back to Kigali, for the selection to be done reliably under the direct control of the central managers – ideally by computer. Implementing the listing operation and the selection of farmers under these conditions may delay the beginning of the first LRSS a few weeks, but it would be worth it.
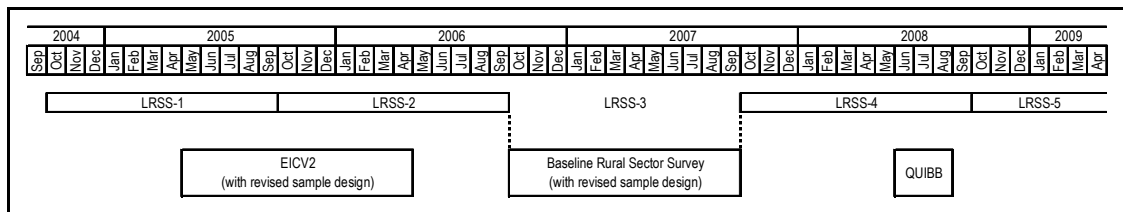
New analytic requirements

The above would have been my only suggestion if the analytic requirements of EICV2 and the rural sector surveys had remained unchanged and if the two of them could still be considered as independent exercises, as was the case in June 2004. These conditions, however, seem to have changed as a result of the requests received from the survey stakeholders during the September 13-14 workshop and subsequent discussions among the participating experts. From the sampling standpoint, the most important conclusion of these discussions is <u>the convenience of having the farmers visited by the BRSS also subjects of the EICV2</u>. A second request was for the EICV2 to revisit some of households observed by the EIVC1 in 2002 (e.g., for incorporating a panel component to the EICV surveys.) The rest of this section explores some of the implications of the new demands on the re-engineering of the sample designs.

A common sample for the EICV2 and the rural sector surveys

Having the EICV2 and the BRSS share a common sample of farmers would permit reducing the farming section of the EICV2 to a minimum, and also correlate the information on crop production collected by the rural sector surveys with the data on consumption and other factors of well-being collected by the EICV2.[6]

Although it is too late to implement this idea for the first round of the LRSS, which is now about to be start (in October 2004,) the EICV2 cannot realistically be fielded before the second quarter of 2005, which gives sufficient time to implement the needed amendments. The second round of the LRSS could then be conducted (starting in October 2005) on a revised sample of farmers, who would <u>not</u> be those visited during the first round, but instead a sub-sample of the EICV2. The third round of the LRSS would be a *de facto* by-product of the BRSS (starting in October 2006). The fourth and subsequent rounds of the LRSS would all be conducted on the panel defined by the second round. The sequence is schematized below:

---

[6] I shall not elaborate further on the reasons for conducting the two surveys over a common sample, since I am sure that this recommendation will be justified much better in the reports of Ms Donovan and Mr McKay, which concentrate on the surveys' subject matters and questionnaires.

2004	2005	2006	2007	2008	2009
Sep Oct Nov Dec	Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec	Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec	Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec	Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec	Jan Feb Mar Apr

LRSS-1	LRSS-2	LRSS-3	LRSS-4	LRSS-5

EICV2 (with revised sample design)

Baseline Rural Sector Survey (with revised sample design)

QUIBB

Clearly, with a common sample design, neither the EICV2 nor the BRSS will be as efficient as they would have been with the *ad hoc* designs proposed last June. However, I think that the two samples have more similarities than differences, and that a common design can be found that only slightly increases the sampling errors of each survey, relative to the June designs. The following preliminary insights may be useful when this is done:

- *Stratification*. In their original designs, the strata of the EICV2 and the BRSS are similar, but not identical. The BRSS is obviously interested in rural areas only, not in the "truly urban" portions of Kigali and a few other cities. Urban areas, on the other hand, are an important concern for the EICV2, to the extent of considering Kigali and the set of all other urban areas in the other eleven provinces as separate strata. The common design may need to define the "truly-urban," "semi-urban" and "rural" portions of all twelve provinces as 36 explicit strata (but not as analytic domains,) and allocate the sample among them in a judicious way that satisfies the requirements of the two surveys well (the "truly urban" part of the sample to be used by the EICV2 only.)

- *First-stage selection probabilities*. In their original designs, the EICV2 selected its PSUs with probability proportional the *total number of households* and the BRSS with probability proportional to the *number of farmers*. The common design will need to arbitrate between the two, and probably will have to opt for the total number of households. The BRSS would not suffer very much from this choice, since most of the households are farmers almost everywhere (except, of course, in the truly urban areas where the BRSS will not be fielded.)

- *Size of the second-stage clusters*. In their original designs, The EICV2 was to select nine households per PSU in urban areas and twelve in rural areas, whereas the BRSS would select twelve households (twelve farmers, actually) per PSU everywhere – even in semi-urban areas. The easiest way of dealing with this would be for the common sample to also take twelve households per PSU in semi-urban areas as well.

- *Second-stage stratification*. In their original designs, neither the EICV2 nor the BRSS implemented stratification at the second stage (although that could have been beneficial for the BRSS, as said before.) In the common design, some kind of second-stage stratification would be unavoidable (at least in rural and semi-urban areas,) since the EICV2 needs to visit all kinds of households and the BRSS farmers only. An option worth considering would be to split the set of all households into three subsets: {a} Irish potato growers, {b} other farmers, and {c} non-farmers. Then, take a sample of twelve farmers for the BRSS from subsets {a} and {b} – with preference for {a} – and call {α} and {β} the resulting sub-samples. Finally, select twelve households for the EICV2 from subsets {α}, {β}

and {c}, with probabilities such that the resulting sample becomes an equal-probability sample of all households in the PSU. (It goes without saying that such an elaborate process could not possibly be entrusted to the field staff; however, it would not be too hard to implement centrally by computer.)

A panel component for the EICVs

Incorporating a panel component to the EICVs could bring about interesting analytic benefits. Having the EICV2 visit some of the same households observed by the EICV1 in 2002 would permit estimating the <u>changes</u> of various living standards indicators more reliably, as well as modelling the <u>reasons</u> of the observed changes better than with two independent cross-sectional samples.[7]

Panel samples, however, make survey planning and management more difficult, and have certain analytic shortcomings, as well as benefits. A serious disadvantage is that a panel, by definition, does not represent the current population of the country, since households created after the previous round have no chance of being observed. This intrinsic selection bias is worsened by *sample attrition* (the practical impossibility of re-interviewing all of the households visited last time,) and, in the specific case of Rwanda, by the fact that the EICV1 used as a sample frame the 1992 Census (which was already old in 2002,) rather than the 2002 Census now available.

Taking both the benefits and the inconveniences of panels into account, and assuming that the wherewithal could be found, my recommendation would be to implement a panel component for the EICV, as a marginal addition, but <u>not as a total or partial replacement</u> of the cross-sectional designs now proposed for the EICV2 (or for the common EICV2/BRSS sample.) One option worthy of consideration would be to append to the EICV2 sample a random sample of about 120 PSUs[8] from the EICV1. The survey could be conducted in the panel component with the same questionnaire and field procedures of the EICV2, perhaps with a few additional questionnaire modules.[9] At the analytical stage, however, the panel and cross-sectional components of the EICV2 should be considered as two separate samples, each representative of a different population.

---

[7] For a more extensive discussion of the benefits of panelling the EICV, see Verwimp, P., *Incorporating a panel component to Rwanda's EICV2 Survey*. Presentation at the EICV2 preparation workshop. Kigali, September 2004.

[8] The exact number of PSUs would depend on the desired proportion of urban-to-rural households in the panel. It should also be consistent with the number of teams added to the EICV2 field effort.

[9] At least one such module would be needed to document the current status of all EICV1 household members, and to record the links between their former and current ID codes.

Appendix 1

Draft plan and budget for the EICV2

EICV2 activities

The GANTT charts below contain a notional plan of activities for the preparation, fielding and analysis of the EICV2.

Tasks have been grouped into eight groups of activities: (1) EICV Financing, management and logistics, (2) EICV Questionnaire development, (3) Sampling, (4) Staffing and training, (5) Fieldwork, (6) Data management, (7) Analysis and dissemination, and (8) Technical assistance.

It was prepared under the assumption that the Rwandan Directorate of Statistics formally becomes an Institute before the end of this year as expected. If this deadline cannot be met, all tasks starting from January 2005 onwards would have to rescheduled accordingly. Another two critical tasks are the pilot testing of the EICV2 questionnaire (task 2.07) and the training of interviewers and supervisors (task 4.09.) Delays in these critical activities would also have an impact on the schedule as a whole.

Time units are weeks. The numbers represent the first Monday of each month.

1	EICV Financing, management and logistics		
1.01	Appoint EICV Project Director	03-Jan-05	09-Jan-05
1.02	Appoint EICV Field Advisor	03-Jan-05	09-Jan-05
1.03	Appoint EICV Data Advisor	03-Jan-05	09-Jan-05
1.04	Finalize institutional agreements and secure financing	03-Jan-05	09-Jan-05
1.05	Secure EICV Core Staff Team premises and logistics	03-Jan-05	09-Jan-05
1.06	Secure premises for data management and storage	03-Jan-05	06-Feb-05
1.07	Acquire data entry computers	03-Jan-05	06-Feb-05
1.08	Acquire anthropometric equipment	03-Jan-05	06-Feb-05
1.09	Acquire vehicles	03-Jan-05	06-Feb-05
1.10	Acquire other survey materials	03-Jan-05	06-Feb-05
1.11	Define household motivation strategy (publicity, etc.)	03-Jan-05	06-Feb-05

2	EICV Questionnaire development		
2.01	First user workshop	13-Sep-04	14-Sep-04
2.02	Summarise user recommendations	15-Sep-04	17-Sep-04
2.03	Prepare and discuss first draft	20-Sep-04	10-Oct-04
2.04	Second user workshop	11-Oct-04	12-Oct-04
2.05	Prepare second draft	13-Oct-04	24-Oct-04
2.06	Prepare pilot test logistics	10-Jan-05	16-Jan-05
2.07	**Pilot test**	17-Jan-05	13-Feb-05
2.08	Review pilot test	14-Feb-05	20-Feb-05
2.09	Finalize questionnaire	14-Feb-05	20-Feb-05
2.10	Translate questionnaire into Kinyarwanda	21-Feb-05	06-Mar-05
2.11	Develop diaries	10-Jan-05	16-Jan-05
2.12	Print questionnaire and diaries	21-Feb-05	20-Mar-05
2.13	Prepare community and price questionnaires	01-Nov-04	16-Jan-05

3	Sampling		
3.01	Select EICV1 Panel PSUs	01-Nov-04	07-Nov-04
3.02	Enter EICV1 Panel HH member names	08-Nov-04	21-Nov-04
3.03	Print EICV1 Panel HH rosters	22-Nov-04	28-Nov-04
3.04	Integrate EICV and Ag Sampling designs	01-Nov-04	07-Nov-04
3.05	Select EICV and Ag PSUs	08-Nov-04	10-Nov-04
3.06	Plan the field assignments	11-Nov-04	13-Nov-04
3.07	Household listing operation	03-Jan-05	03-Apr-05
3.08	Get hh lists and select hh's	10-Jan-05	10-Apr-05
3.09	Estimate sampling errors of key variables	15-May-06	01-Jun-06

4 Staffing and training

4.01	Select and train pilot test staff	10-Jan-05	16-Jan-05
4.02	Prepare field manuals	31-Jan-05	06-Feb-05
4.03	Prepare quality control procedures and manuals	31-Jan-05	13-Feb-05
4.04	Select candidates for interviewer/supervisor training	10-Jan-05	27-Mar-05
4.05	Identify anthropometrists	10-Jan-05	27-Mar-05
4.06	Select PSUs fror training practice	14-Mar-05	27-Mar-05
4.07	Prepare training venue and logistics	21-Mar-05	27-Mar-05
4.08	Train master trainers	21-Mar-05	03-Apr-05
4.09	**Train interviewers and supervisors**	**04-Apr-05**	**01-May-05**
4.10	Train anthropometrists	18-Apr-05	01-May-05

5 Field work

5.01	First RSLS in the field	04-Oct-04	02-Oct-05
5.02	EICV in the field	02-May-05	30-Apr-06
5.03	Agrculture Baseline survey in the field	01-Oct-06	30-Sep-07

6 Data management

6.01	Prepare First Version of Data Entry Program	03-Jan-05	30-Jan-05
6.02	Debug Data Entry Program with pilot test data	31-Jan-05	13-Feb-05
6.03	Finalize Data Entry Program	14-Feb-05	27-Mar-05
6.04	Select candidates for Data Entry Operator training	10-Jan-05	27-Mar-05
6.05	Install Data Entry computers	28-Mar-05	03-Apr-05
6.06	Train Data Entry operators	04-Apr-05	01-May-05
6.07	Data Entry (month 1)	02-May-05	05-Jun-05
6.08	Evaluation and debugging	06-Jun-05	19-Jun-05
6.09	Data Entry (months 2-12)	06-Jun-05	30-Apr-06

7 Analysis and dissemination

7.01	Define contents of basic reports	10-Sep-05	07-Oct-05
7.02	Create datasets for first six months	21-Nov-05	04-Dec-05
7.03	Prepare preliminary report	05-Dec-05	05-Feb-06
7.04	Distribute preliminary report	06-Feb-06	05-Mar-06
7.05	Seminar	06-Mar-06	12-Mar-06
7.06	Revise contents of preliminary report	01-May-06	14-May-06
7.07	Create final datasets	15-May-06	04-Jun-06
7.08	Prepare statistical summaries	05-Jun-06	25-Jun-06
7.09	Document datasets	13-Mar-06	30-Apr-06
7.10	Disseminate datasets	26-Jun-06	01-Jul-06

8 Technical Assistance

8.01	Field Advisor	01-Jan-05	31-Dec-06
8.02	Data Advisor	01-Jan-05	31-Dec-06
8.03	Questionnaire development	04-Oct-04	17-Oct-04
8.04	Finalize sampling design	01-Nov-04	14-Nov-04
8.05	Pilot testing	17-Jan-05	20-Feb-05
8.06	Training of anthropometrists	18-Apr-05	01-May-05
8.07	Data Management I	03-Jan-05	23-Jan-05
8.08	Survey start-up	25-Apr-05	15-May-05
8.09	Data Management II	18-Apr-05	08-May-05
8.10	Data Management III	28-Nov-05	11-Dec-05
8.11	Mid-term assessment	28-Nov-05	11-Dec-05
8.12	Data analysis I	05-Dec-05	25-Dec-05
8.13	Data analysis II (Poverty Profile)	01-Jun-06	31-Jul-06
8.14	Data analysis III (Revised Poverty Map)	05-Jun-06	02-Jul-06
8.15	Estimate sampling errors	05-Jun-06	25-Jun-06

Appendix 2

Workshop presentations

This appendix reproduces the slides of the two presentations offered at the September 14 workshop with stakeholders.

The first presentation, on the sampling options available to the EICVs, started by giving some of the basic facts of survey sampling, to focus later on some of the specific challenges faced by the EICV2, namely the importance of randomizing the schedule for visiting the selected PSUs within the twelve-month period of field operations, and the tradeoffs of endowing it with a panel component.

EICV sampling issues

Choices and pitfalls
Allocating the sample in space and time
Monitoring indicators
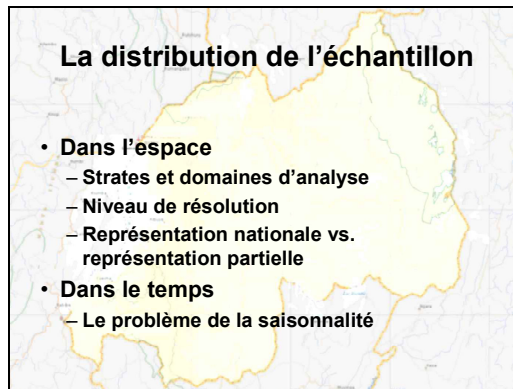(benefits and disadvantages of panel surveys)
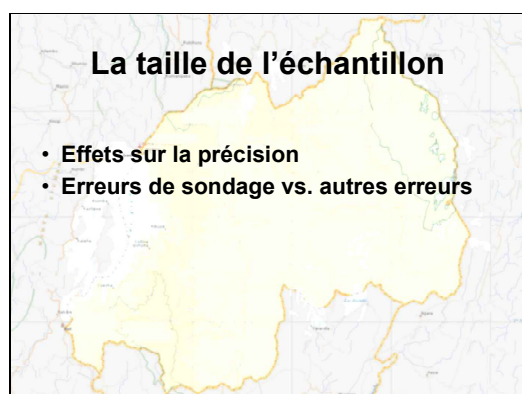
The size of the sample:
its effects on precision - sampling vs. non-sampling errors
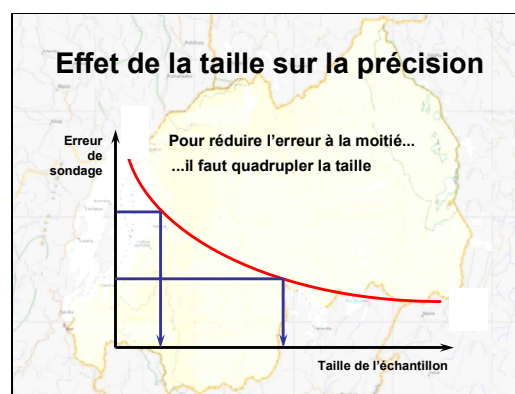
Effect of sample size on precision:
To halve sampling errors…
Sample size must be quadrupled

Erreurs de sondage vs. autres erreurs

Sampling vs. non-sampling errors

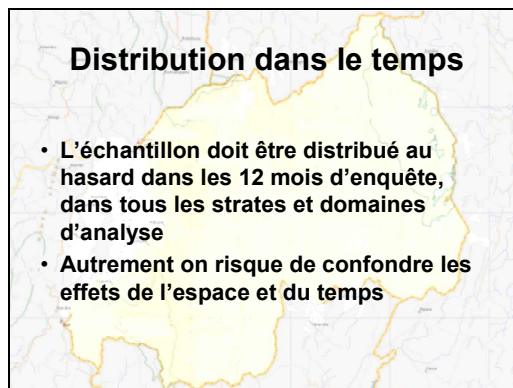### La distribution de l'échantillon
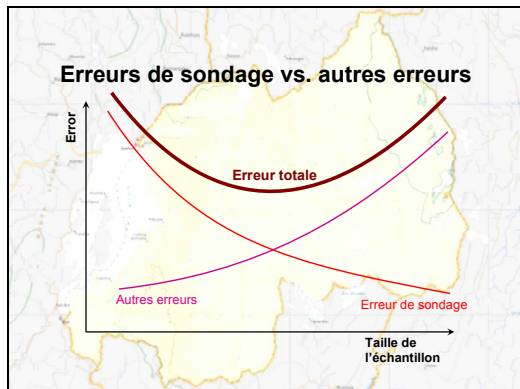
- **Dans l'espace**
  - **Strates et domaines d'analyse**
  - **Niveau de résolution**
  - **Représentation nationale vs. représentation partielle**
- **Dans le temps**
  - **Le problème de la saisonnalité**

Allocation of the sample
In space (strata and analytic domains – resolution level – national vs. sub-national representativity)
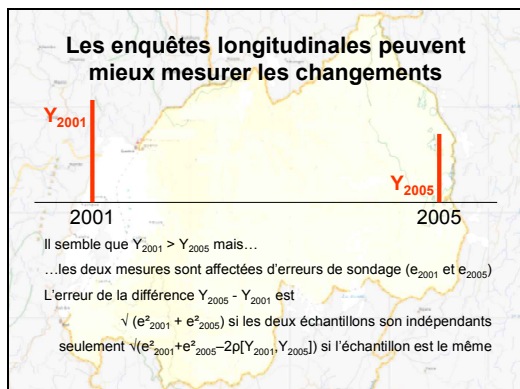In time (the problem of seasonality)

Distribution dans l'espace

	UPSs	Ménages par UPS	Ménages 2001	Ménages 2005
Urbain				
Kigali Ville	80	9	720	900
Other Urbain	50	9	450	720
Total urbain	**130**		**1,170**	**1,620**
Rural				
Kigali ngali	40	12	480	480
Gitarama	40	12	480	480
Butare	40	12	480	480
Gikongoro	40	12	480	480
Cyangugu	40	12	480	480
Kibuye	40	12	480	480
Gisenyi	40	12	480	480
Ruhengeri	40	12	480	480
Byumba	40	12	480	480
Umutara	40	12	480	480
Kibungo	40	12	480	480
Total rural	**440**		**5,280**	**5,280**
RWANDA	**570**		**6,450**	**6,900**

Les pièges:

Tous les ménages Rwandais n'ont pas la même chance d'être visités par l'EICV:

Il faut pondérer les résultats !

Le sondage est à deux degrés (pas un sondage aléatoire simple):

Les erreurs de sondage sont affectés par l'effet de grappe !

La définition des secteurs urbains a changé entre 2001 et 2005:

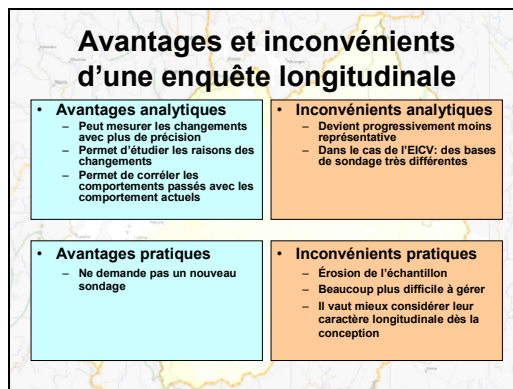**Il faut comparer les enquêtes avec beaucoup de précaution !**

Allocation in space – Pitfalls:
All Rwandan households don't have the same chances of being visited by the EICV: results must be weighed
Sampling is in two stages: must consider cluster effects
Definition of urban changed: 2001 and 2005 results must be compared with caution

Distribution dans le temps

- **L'échantillon doit être distribué au hasard dans les 12 mois d'enquête, dans tous les strates et domaines d'analyse**
- **Autrement on risque de confondre les effets de l'espace et du temps**

Allocation in time:
Sample must randomly distributed among the 12-month data collection period, within all strata and analytic domains.
Time and space risk can be confused otherwise.

Les enquêtes longitudinales peuvent mieux mesurer les changements

Il semble que $Y_{2001} > Y_{2005}$ mais…

…les deux mesures sont affectées d'erreurs de sondage (e_{2001} et e_{2005})

L'erreur de la différence $Y_{2005} - Y_{2001}$ est

$\sqrt{(e^2_{2001} + e^2_{2005})}$ si les deux échantillons son indépendants

seulement $\sqrt{(e^2_{2001}+e^2_{2005}-2\rho[Y_{2001},Y_{2005}])}$ si l'échantillon est le même

Panel samples can measure changes better
It seems that $Y_{2001} > Y_{2005}$ but… both measures are affected by sampling errors (e_{2001} and e_{2005})
The error of the difference is
$\sqrt{(e^2_{2001} + e^2_{2005})}$ if the samples are independent
only $\sqrt{(e^2_{2001}+e^2_{2005}-2\rho[Y_{2001},Y_{2005}])}$ if the sample is the same

Avantages et inconvénients d'une enquête longitudinale

- **Avantages analytiques**
  - Peut mesurer les changements avec plus de précision
  - Permet d'étudier les raisons des changements
  - Permet de corréler les comportements passés avec les comportement actuels

- **Inconvénients analytiques**
  - Devient progressivement moins représentative
  - Dans le cas de l'EICV: des bases de sondage très différentes

- **Avantages pratiques**
  - Ne demande pas un nouveau sondage

- **Inconvénients pratiques**
  - Érosion de l'échantillon
  - Beaucoup plus difficile à gérer
  - Il vaut mieux considérer leur caractère longitudinale dès la conception

Advantages and disadvantages of panel surveys
Analytic advantages: measure changes better; study the reason of changes; correlate past and present behaviours.
Analytic disadvantages: the sample becomes progressively less representative; in the case of the EICV, also very different sample frames.
Practical disadvantages: sample erosion; much harder management; better to think about panelling from the beginning.
Practical advantages: no need to sample again.

The second presentation, on survey data management, exposed the advantages of integrating computer-based quality controls to survey field operations, emphasizing the importance of at least conserving for EICV2 the parallel techniques and procedures implemented successfully by the EICV1 in 2001.

La qualité des enquêtes

**Le produit finale de l'enquête est une
base des données**

**Comment pouvons-nous assurer la qualité
du produit fourni à nos clients?**

The quality of surveys
The end product of a survey is a **database**
How can we deliver to our clients a quality product?

L'informatique et les opérations de terrain

- Que se passe-t-il quand on ne les intègre pas?
- Bénéfices de l'intégration
- Options tactiques
- Aspects stratégiques

Data management and field operations
What happens when they are not integrated?: Benefits of integration - Tactic options - Strategic considerations

Que se passe-t-il quand on ne les intègre pas?

On impose l'étape d'apurement,
qui peut être longe et frustrante

Les données peuvent perdre sa
relevance pour la prise de décisions

La qualité des données n'est plus garantie

Dans le meilleure des cas l'apurement
conduit à une base de données
internement consistante

L'apurement entraîne innombrables décisions,
souvent mal documentées

Les utilisateurs se méfient

What happens when data management and field operations are not integrated?
A long and frustrating "data cleaning" phase becomes unavoidable: the data can loose its policy relevance
Data quality cannot be guaranteed: data cleaning at best converges to *internally consistent* datasets.
Data cleaning entails a myriad of generally undocumented decisions: users mistrust.

Bénéfices de l'intégration

- Fournit des bases de données **fiables** en temps **opportun**
- Fournit un feedback immédiat sur les performances des équipes de terrain, permettant ainsi la **détection précoce des comportements inadéquats**
- Assure l'application des **critères uniformes** pendant toute la période de collecte, et de la part de tous le personnel de terrain
- Résout les inconsistances par vérification directe de la réalité **auprès des ménages**, plutôt que par des réflexions de bureau
- Est consistante avec la culture de **qualité totale**

Benefits of integration
Provides **reliable** databases on a **timely** basis.
Provides immediate feedback on the performance of field teams, thus allowing **early detection** of inadequacies.
Fosters the application of **uniform criteria** by all field staff during the whole data collection period.
Solves inconsistencies by actually re-visiting **the households** rather than through office guesswork.
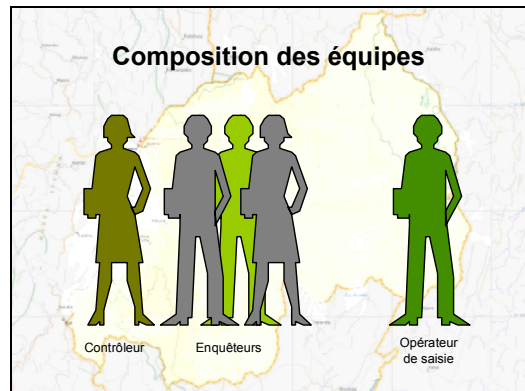Is consistent with the **total quality** paradigms.

Tactic options:
Parallel data entry – Mobile teams with fixed data entry – mobile data entry –
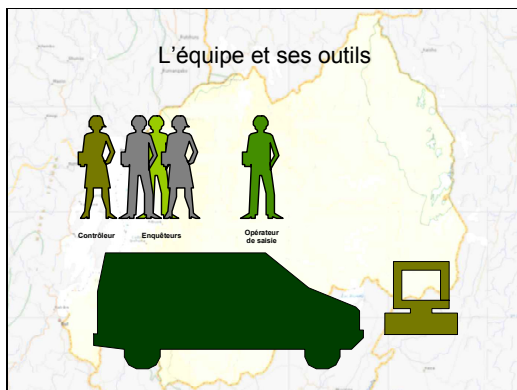Paperless interviewing
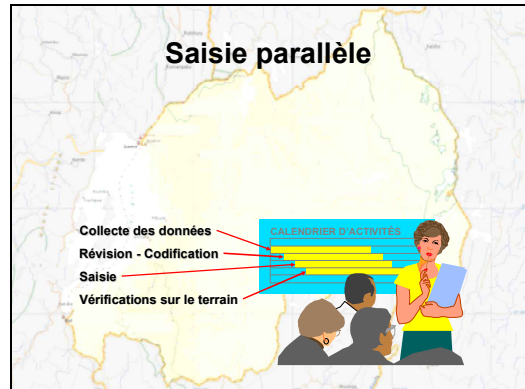
Parallel data entry
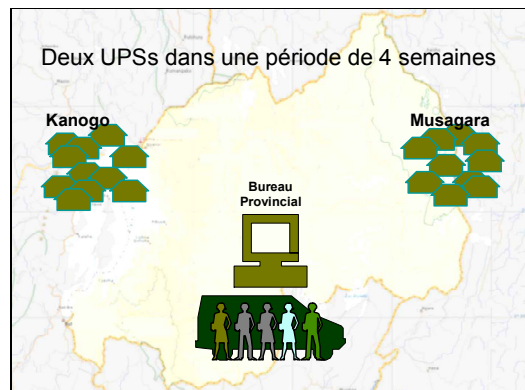Data collection – Checking and Coding – Data Entry – Field verifications
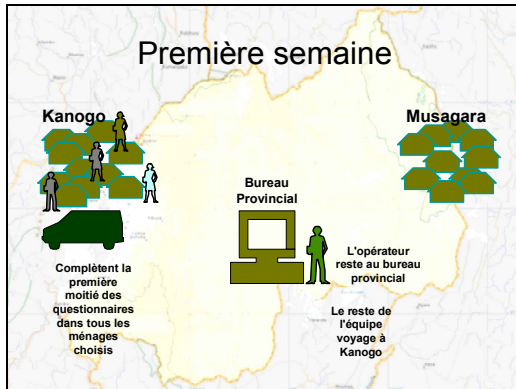

Mobile teams with fixed data entry

Team composition
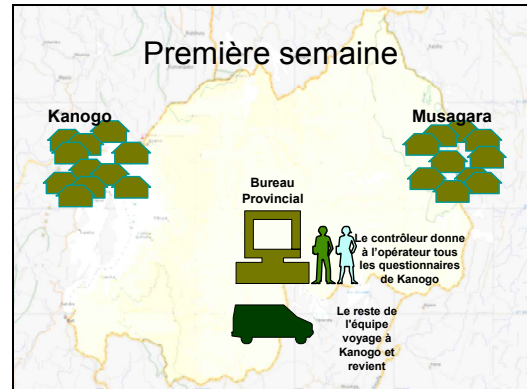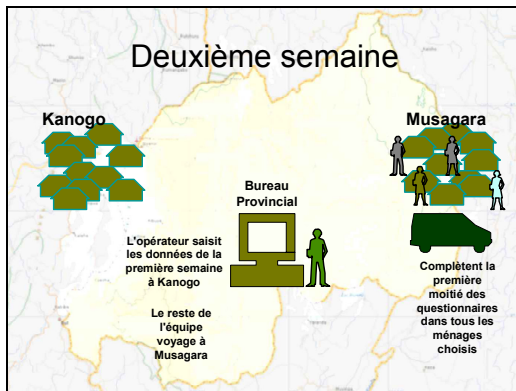Supervisor – Interviewers – Data Entry Operator (DEO)


The team and its tools

Two PSUs in a 4-week period

First week:
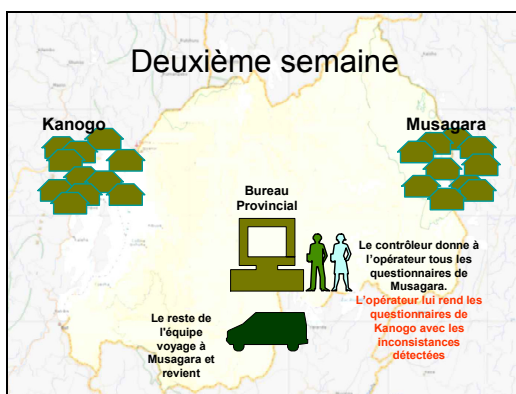The team completes the first half of all questionnaires in Kanogo – The DEO stays in the provincial office.

First week:
The team returns from Kanogo - The supervisor gives the Kanogo questionnaires to the DEO.

Second week:
The DEO enters the first half of all questionnaires from Kanogo – The rest of the team travels to Musagara to complete the first half of the questionnaires.
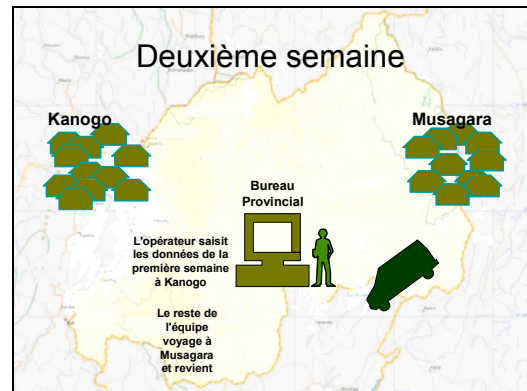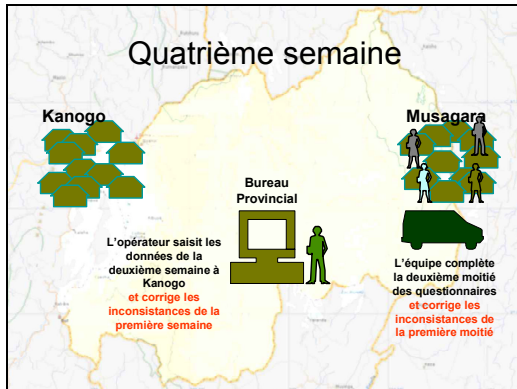
Second week:
The rest of the team returns from Musagara.

Second week:
The supervisor gives the Musagara questionnaires to the DEO – The DEO gives the supervisor the Kanogo questionnaires with flagged inconsistencies.
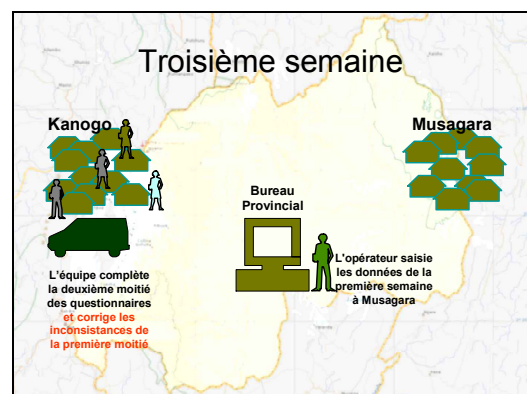
Third week:
The team completes the second half of the Kanogo questionnaires and corrects the inconsistencies flagged for the first half – The DEO enters the first half of the Musgara questionnaires.
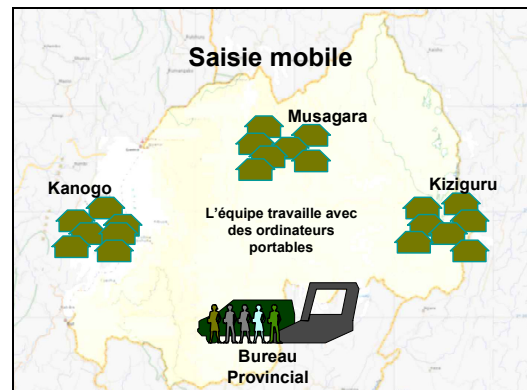
Fourth week:
The team completes the second half of the Musagara questionnaires – The DEO enters the second half of the Kanogo questionnaires and corrects the inconsistencies of the first half.
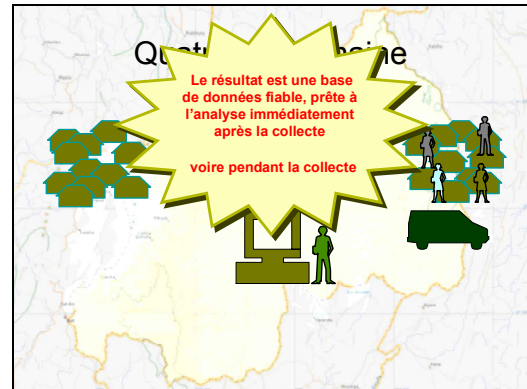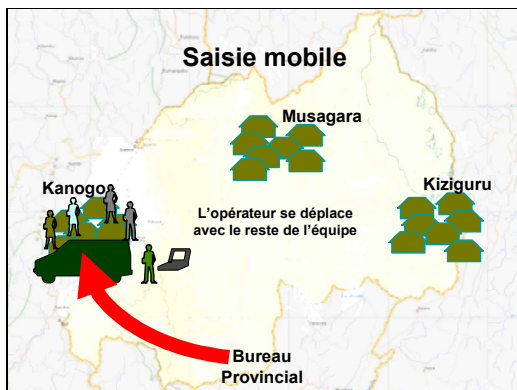

The result is a reliable dataset, ready for analysis immediately after data collection
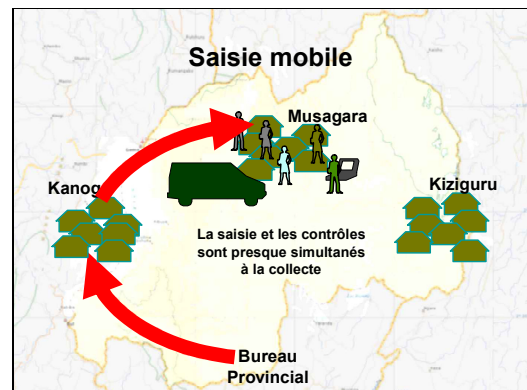– even during data collection.
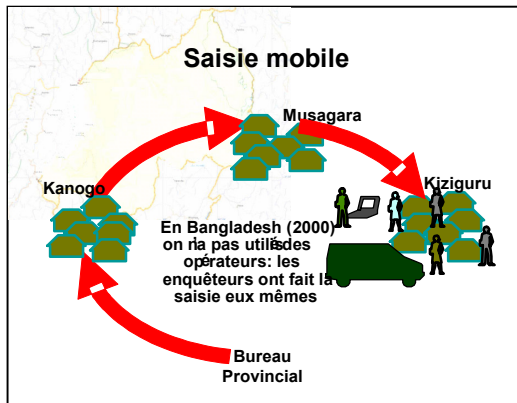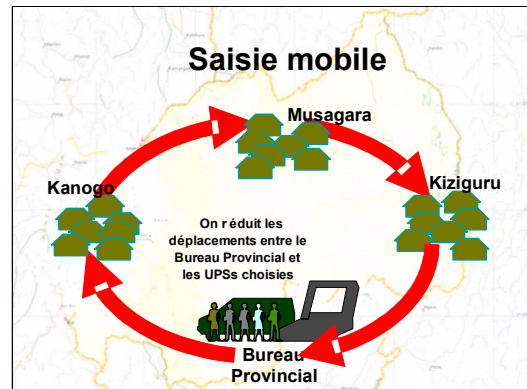

Mobile data entry

The team works with laptops

The operator travels with the rest of the team

Data entry and quality controls are almost concurrent with data collection

Saisie mobile

Musagara

Kanogo — Kiziguru

En Bangladesh (2000) on n'a pas utilisé des opérateurs: les enquêteurs ont fait la saisie eux mêmes

Bureau Provincial

In Bangladesh (2000) no DEOs were used: interviewers entered the data themselves.

Saisie mobile

Musagara

Kanogo — Kiziguru

On réduit les déplacements entre le Bureau Provincial et les UPSs choisies

Bureau Provincial

Reduced travel to and from provincial offices

Interviews sans papier

- C'est l'option du future
- Appliquée avec succès dans des enquêtes plus simples
- Le technologie est disponible, mais il faut encore résoudre
  - La conception du questionnaire
  - L'ergonomie des interviews
  - La formation des enquêteurs
  - La supervision

Paperless interview
It's the option of the future
Has been applied successfully to simple surveys
Technology is available, but details need to be solved: questionnaire design - ergonomics of the interview – interviewer training – supervision.

Considérations stratégiques

L'intégration de l'informatique aux opérations de terrain

- Exige un planning très soigné de toutes les étapes de la conception de l'enquête
- Entraîne l'intégration des dextérités et la collaboration
  - sur le terrain et
  - au niveau de la gestion centrale du projet
- Est un complément, pas un substitut des schémas de formation et supervision
- Elle peut cependant réduire les effets nocifs de la formation ou de la supervision inadéquates

Strategic considerations
The integration of computer-based quality controls:
Requires careful planning of all survey design phases;
Involves integration of skills and collaboration in the field and in central management;
Is a complement, not a replacement of good survey training and supervision
Can reduce, however, the ill effects of inadequate training and supervision.

La formation des enquêteurs

- Commence avec la sélection
  - Credentials, références, expérience
  - Aptitudes pour le travail de terrain
- Demande la formation des formateurs
- Demande une préparation très soignée
  - Aspects logistiques
  - Moyens audio-visuels
  - Ménages pour pratiquer
- Trois éléments
  - Séances pléniers
  - Séances de groupe (~ 20 enquêteurs par groupe)
  - Pratiques auprès des ménages
- Finit avec la sélection

Interviewer training
Begins with preliminary selection: credentials, references, experience – aptitudes for fieldwork
Requires training of trainers
Requires careful preparation: logistics – audio-visual tools – households for practice
Three elements: plenary sessions – group sessions – field practice
Ends with final selection

Les tâches du contrôleur

- **Administration**
  - **Suivie du travail des enquêteurs**
  - **Gestion des questionnaires**
  - **Gestion du budget, des moyens de transport, etc.**
  - **Remplacer les enquêteurs en cas de besoin**
  - **Etc.**
- **Contrôle de qualité**
  - **Formation permanente des enquêteurs**
  - **Vérification visuelle des questionnaires**
  - **Observation des interviews**
  - **Visites de contrôle auprès des ménages**
  - **Supervision de la saisie**

The role of the supervisor
Management: monitoring the work of interviewers – questionnaire management – managing budget, means of transportation, etc. – replacing interviewers – Etc.
Quality controls: ongoing training of interviewers – visual inspection of questionnaires – observation of interviews – check-up visits to households – supervision of data entry