Techniques d'enquête, juin 1995
Vol. 21, n° 1, pp. 99–107
Statistique Canada



# Enquêtes par sondage auprès des visiteurs

RONALDO IACHAN et SUZANNE S. KEMP[1]

RÉSUMÉ

Dans le présent article, on traite de la conception d'enquêtes auprès des visiteurs. Pour illustrer le sujet, on décrit deux enquêtes récentes. La première est une enquête auprès des visiteurs des zones administrées par le National Park Service à l'échelle des États-Unis pendant toute une année (1992). La deuxième est une enquête menée auprès des personnes qui ont utilisé des installations de loisirs du bassin des trois rivières autour de Pittsburgh, en Pennsylvanie, pendant une période de douze mois. Les deux enquêtes comportaient un échantillonnage dans le temps avec stratification temporelle et stratification dans l'espace. Les unités d'échantillonnage avaient la forme de paires site-période à l'avant dernier degré d'échantillonnage soit celui précédant de l'échantillonnage des visiteurs. L'attribution aléatoire des sites de l'échantillon à des périodes permet d'obtenir des estimations non biaisées pour les strates temporelles (p. ex., des estimations mensuelles et saisonnières) ainsi que des estimations pour les strates définies selon la région et le genre d'utilisation.

MOTS CLÉS: Personne qui utilise des installations de loisirs; échantillonnage dans le temps; site-période.

## 1. INTRODUCTION

Les enquêtes auprès des visiteurs présentent des défis particuliers dont on traite rarement dans les publications statistiques. Le présent article tente de combler cette lacune en décrivant la conception de deux enquêtes réalisées récemment par le Research Triangle Institute (RTI) et en mettant l'accent sur les caractéristiques communes de ces deux enquêtes. Nous espérons que les leçons apprises au cours de ces travaux profiteront aux chercheurs qui planifient des enquêtes semblables.

La première enquête était une étude des visiteurs des zones administrées par le National Park Service (NPS), réalisée conjointement pour le National Park Service par RTI et HBRS, Inc. Cette étude comportait la production d'un échantillon probabiliste de visiteurs des parcs, qui représentait les visiteurs de 323 zones administrées par le NPS dans tous les États-Unis (à l'exception de l'Alaska) pendant toute l'année 1992. Pour simplifier la rédaction, nous désignerons les zones administrées par le NPS par l'expression "parcs", tout en signalant que ces zones comprennent aussi des sites historiques ainsi que des parcs à vocation culturelle. Le principal objectif de l'enquête du NPS était d'évaluer les expériences et les problèmes des visiteurs en portant une attention particulière aux problèmes liés au survol des parcs par des avions (p. ex., le bruit et autres gênes possibles). Diverses données ont aussi été recueillies au moyen d'une enquête postale portant sur un sous-échantillon des visiteurs sélectionnés.

La deuxième enquête était une étude portant sur les personnes qui utilisent des installations de loisirs du bassin des trois rivières, dans la région de Pittsburgh, le long des rivières Monongahela, Allegheny et Ohio en 1992 (ou, plus précisément, entre février 1992 et janvier 1993). Cette enquête a été réalisée conjointement pour la Ohio River Valley Sanitation Commission par RTI et Terrestrial Environmental Specialists. La région visée par l'étude comprenait un segment de 40 milles de la rivière Ohio, un segment de 24 milles de la rivière Monongahela ainsi qu'un segment de 7 milles de la rivière Allegheny. L'objectif principal de l'étude sur les trois rivières était de construire un profil de base des utilisateurs des installations de loisirs de la région et de modéliser la valeur économique que ces derniers attribuent à diverses activités. On a fait la distinction entre trois types de base d'activités de loisirs et de sports: navigation de plaisance, pêche sportive et fréquentation des parcs.

L'étude portant sur les trois rivières est la plus complète d'une série d'études réalisées par RTI pour évaluer l'impact sur l'environnement dans un certain nombre d'États. Ces études estiment la réduction possible de la valeur économique ou récréative attribuée par les personnes qui utilisent effectivement ou qui pourraient utiliser des installations de loisirs à des zones qui ont été ou qui pourraient être touchées. Bien que, pour une enquête plus étendue portant sur les utilisateurs éventuels de telles zones, on puisse considérer un plan d'échantillonnage dans le cadre duquel les répondants sont interviewés par téléphone, on a trouvé qu'il fallait avoir recours à des interviews sur le vif auprès des visiteurs afin d'obtenir de ces derniers des renseignements à un moment très rapproché de celui où ils utilisent effectivement les installations.

Une discussion des problèmes liés à la conception d'enquêtes auprès des visiteurs comme celles qui viennent d'être mentionnées a été présentée récemment dans Kalton (1991), notamment en ce qui a trait à l'échantillonnage dans le temps et dans l'espace dont l'importance est critique dans notre cadre de travail. Sous sa forme la plus simple, un plan d'échantillonnage prototype à deux degrés pour une enquête auprès des visiteurs considère des paires site-période

[1] Ronaldo Iachan, Statisticien chercheur et Suzanne S. Kemp, Statisticienne, Research Triangle Institute, 3040 Cornwallis Road, Research Triangle Park, NC 27709-2194, U.S.A.

comme unités primaires d'échantillonnage (UPÉ) à partir desquelles les visiteurs sont choisis au deuxième degré. Les exemples comprennent les sondages faits à la sortie de l'isoloir (voir, par exemple, Levy 1983), les interviews sur le vif dans les centres commerciaux (voir, par exemple, Sudman 1980) et d'autres enquêtes sur les transports et la circulation (Gough et Ghangurde 1977; Kish, Lovejoy et Rackow 1961).

Parmi les questions relatives au plan d'échantillonnage qui sont importantes dans les enquêtes auprès des visiteurs, on peut distinguer les problèmes généraux suivants:

- Il est souhaitable de choisir avec une plus grande probabilité les paires site-période pour lesquelles il y a un plus grand nombre de visiteurs; la stratification et le choix avec PPT sont alors des caractéristiques efficaces du plan.
- Les paramètres relatifs à la collecte des données sont la clé pour préciser la longueur de la période ainsi que les taux d'échantillonnage pour chaque paire site-période; p. ex., on doit faire certains compromis entre la possibilité que les employés sur le terrain soient trop occupés (périodes brèves, taux d'échantillonnage élevés) ou pas assez occupés (longues périodes, faibles taux d'échantillonnage).
- Les objectifs sur le plan de l'analyse ainsi que l'efficience nous incitent à utiliser, pour la stratification temporelle, des dimensions telles que la saison, le mois, la fin de semaine par opposition au jour de semaine, et même à la période de la journée; p. ex., le besoin d'estimations saisonnières laisse supposer l'utilisation de saisons ou de mois comme strates pour le choix des périodes.

Les deux études dont on traite dans cet article ont un objectif principal commun, soit celui de caractériser la population des visiteurs dans la région ou dans le pays sur une année complète. Toutefois, dans chaque cas, elles diffèrent au niveau des estimations les plus importantes qui mènent aux caractéristiques fondamentales du plan. Dans l'étude sur les trois rivières, mais non dans l'étude du NPS, on devait calculer des estimations mensuelles assez précises. Pour la première étude, alors, les unités d'échantillonnage dans le temps, en l'occurrence les jours, ont été stratifiées en mois. La stratification dans l'espace des sites de l'étude sur les trois rivières était géographique et par type d'activités de loisirs (navigation de plaisance, pêche sportive ou fréquentation des parcs).

Pour l'étude du NPS, la stratification primaire a été faite selon le type de parc. Certains parcs devaient absolument être inclus dans l'échantillon pour répondre à des prescriptions de la loi. Dans ces parcs et dans d'autres parcs choisis, on souhaitait aussi calculer des estimations propres à chaque parc. Pour ces parcs, dits parcs très achalandés, on a alors décidé de choisir relativement plus de paires site-période. Le problème d'optimisation pour le plan d'échantillonnage initial était de savoir comment répartir la taille de l'échantillon entre les degrés d'échantillonnage, c.-à-d. de déterminer combien de parcs et combien de paires site-période par parc devraient être choisis. Dans la section 3, on étudie une solution à ce problème qui est une fonction des corrélations à l'intérieur d'une grappe dans les parcs et dans les périodes. L'optimisation du plan pour l'enquête du NPS s'appliquait aussi aux strates temporelles et spatiales aux degrés intermédiaires d'échantillonnage, entre les parcs pour le premier degré et les paires site-période pour l'avant-dernier degré (en n'oubliant pas que les visiteurs sont choisis au dernier degré).

Ces deux enquêtes illustrent des questions telles que la stratification temporelle, le choix des unités d'échantillonnage appropriées et l'attribution aléatoire d'unités géographiques à des unités temporelles. Dans la section 2, on présente les aspects communs aux deux enquêtes ainsi que leurs différences fondamentales. Dans les sections 3 et 4, on décrit le plan de l'enquête auprès des visiteurs du NPS ainsi que celui de l'enquête sur les trois rivières, respectivement. Dans la section 5, on traite des procédures de pondération utilisées pour les enquêtes. Un bref aperçu et certaines conclusions sont présentés à la section 6.

## 2. APERÇU DES PLANS D'ÉCHANTILLONNAGE: POINTS COMMUNS ET DIFFÉRENCES

Pour les deux enquêtes, les derniers échantillons de visiteurs étaient choisis au moyen d'interviews sur le vif au moment où les visiteurs quittaient les lieux choisis dans l'échantillon pendant les périodes sélectionnées. Il fallait effectuer des interviews à la sortie afin que l'on puisse relever l'attitude des visiteurs immédiatement après qu'ils aient utilisé les installations de loisirs. De plus, dans les deux études, les visiteurs étaient choisis à partir d'échantillons de paires site-période. L'utilisation de paires site-période comme unités d'échantillonnage remonte à l'étude de Kish, Lovejoy et Rackow (1961). Cette définition des unités d'échantillonnage permet de choisir les visiteurs selon un calendrier de collecte des données qui précise quels sites seront visés à quel moment. Contrairement à l'enquête portant sur les trois rivières, le choix des paires site-période ne constituait pas le premier degré pour l'enquête du NPS. Les unités primaires d'échantillonnage (UPÉ) pour cette enquête étaient les zones du NPS, ou les parcs. L'enquête du NPS comprenait plusieurs degrés de sélection, qui sont décrits dans la prochaine section.

De plus, les deux enquêtes utilisaient des bases temporelles formées de jours et des périodes de collecte des données admissibles au cours des journées, afin de permettre de faire des inférences pour toute l'année. Les plans d'échantillonnage incluaient le choix de périodes afin que chaque période admissible ait une probabilité positive connue d'être choisie. Si les deux enquêtes faisaient appel à des bases temporelles, la structure des bases et le choix des jours dans chaque enquête étaient, eux, fort différents.

Par exemple, pour l'étude portant sur les trois rivières, l'échantillon choisi était constitué de douze échantillons mensuels indépendants. Chaque échantillon mensuel avait essentiellement le même plan d'échantillonnage aléatoire stratifié, mais on utilisait une répartition différente de l'échantillon ainsi que des tailles d'échantillon différentes au cours des différents mois. Ce plan tenait compte des variations saisonnières dans les caractéristiques des loisirs et des sports et permettait le calcul d'estimations pour

chaque mois et pour chaque strate (p. ex., selon le genre: navigation de plaisance, pêche sportive ou fréquentation des parcs). On permettait aux bases de sondage spatiales et temporelles de varier d'un mois à l'autre. On traite de la stratification ainsi que de la répartition pour cet échantillon dans la section 4.

Par opposition à la situation qui vient d'être décrite, dans la base de sondage temporelle de l'enquête auprès des visiteurs du NPS, on considérait tout d'abord des blocs de deux mois pour chaque parc dans l'échantillon (UPÉ). L'utilisation de périodes de deux mois comme unités d'échantillonnage (au deuxième degré) dans le temps permettait d'atteindre de façon efficace les objectifs de l'enquête pour deux raisons fondamentales. Premièrement, elle permettait une concentration (géographique) efficace des ressources en personnel et répartissait la collecte des données sur toute l'année. Deuxièmement, ce choix de périodes permettait de saisir les fluctuations saisonnières des visites dans tout le réseau des parcs, fluctuations qui découlent du fait que, dans certains parcs, le nombre de visiteurs est relativement plus élevé au printemps alors que pour d'autres, il est plus élevé en automne et ainsi de suite.

Un bloc de deux mois a été choisi pour chaque parc dans l'échantillon afin que la collecte des données puisse être concentrée dans le temps de façon efficace. Puis, à l'étape suivante de la sélection dans le temps, des jours ont été choisis dans le bloc de deux mois pour chaque parc dans l'échantillon. Comme les parcs eux-mêmes, au premier degré, ces blocs de deux mois ont été choisis avec probabilités proportionnelles à la taille (PPT), la mesure de taille étant l'ensemble des visiteurs.

On a équilibré avec soin la taille de l'échantillon ainsi que sa répartition entre les différents degrés d'échantillonnage afin de minimiser les effets de groupement associés aux grappes dans le temps et dans l'espace. Pour l'étude portant sur les trois rivières, ce groupement se produit au premier degré de la sélection, quand les unités d'échantillonnage sont les sites et les périodes. Dans la répartition, on a aussi tenu compte du fait que la taille des échantillons varie dans des échantillons mensuels indépendants successifs. Pour l'enquête du NPS, les grappes temporelles provenaient des blocs de deux mois et des jours choisis aux différents degrés. Les grappes géographiques provenaient de l'utilisation des parcs et des sorties de parcs comme unités d'échantillonnage pour cette enquête.

Dans la section ci-après, on décrit plus en détail le plan utilisé pour l'enquête du NPS.

## 3. PLAN D'ÉCHANTILLONNAGE DE L'ENQUÊTE DU NPS

Le plan d'échantillonnage de l'enquête auprès des visiteurs du NPS tirait profit d'informations supplémentaires de divers genres et provenant de différentes sources:

- Renseignements obtenus lors d'études antérieures (p. ex., le classement des parcs fondé sur l'exposition au bruit ainsi que des classifications faites par des employés du NPS).
- Renseignements disponibles dans les bases de données du NPS (p. ex., les données sur les visiteurs des parcs selon le mois).
- Renseignements recueillis par les employés du NPS expressément pour établir le plan d'échantillonnage (p. ex., un répertoire des sorties de chaque parc dans l'échantillon ainsi que du nombre de véhicules qui quittent le parc par chacune des sorties).

Dans les sous-sections ci-après, on décrit les divers degrés de la sélection pour les interviews sur le vif auprès des visiteurs. Cette partie de l'enquête sera aussi désignée comme l'enquête sur l'avant-pays pour la distinguer d'une enquête sur les utilisateurs des installations de l'arrière-pays, qui a été réalisée en même temps dans les parcs échantillonnés. Dans la sous-section 3.5, on décrit l'enquête sur l'utilisation des installations de l'arrière-pays ainsi qu'une enquête postale réalisée auprès d'un sous-échantillon de répondants à l'enquête sur l'avant-pays ainsi qu'auprès d'un échantillon d'utilisateurs des installations de l'arrière-pays.

Le quatrième degré de la sélection, pour l'enquête postale, comprenait des degrés (et des phases) additionnels de sélection. Pour l'enquête comportant des interviews sur le vif auprès des visiteurs, des groupes (c.-à-d. des véhicules) étaient choisis comme grappe ultime: on demandait à toutes les personnes dans un groupe dans l'échantillon si elles acceptaient d'être interviewées. Pour l'enquête postale, on a tiré un sous-échantillon dans les groupes et une personne a été choisie dans chaque groupe faisant partie du sous-échantillon.

### 3.1 Construction de la base de sondage et premier degré d'échantillonnage

Nous avons construit une base de sondage pour le choix des parcs en compilant des renseignements du NPS sur les visites (mensuelles et annuelles) dans les parcs et sur l'exposition au bruit, renseignements qui étaient utilisés de deux façons distinctes dans le plan d'échantillonnage: pour la stratification et pour l'attribution des mesures de taille. Ces derniers renseignements étaient basés sur deux sources différentes: a) une étude antérieure du NPS dans laquelle les parcs étaient classés en fonction de l'exposition potentielle et b) une classification des parcs effectuée indépendamment par des employés du NPS (surintendants de parc, employés travaillant dans les régions, *etc.*).

En consultation avec le NPS, RTI a combiné ces deux classifications pour l'exposition au bruit afin de construire onze strates. La stratification répartit les parcs en catégories très élevée, élevée, faible et très faible. Les strates ont été divisées en deux sous-strates à l'aide du classement dans la strate. (Il est à noter que la strate "moyenne" n'a pas été subdivisée à cause du faible nombre de parcs qu'elle contenait.)

En plus de l'exposition au bruit, ces strates incorporaient trois classes de parcs qui méritent un traitement distinct: 1) parcs urbains et de banlieue, 2) parcs pour lesquels des données sur les visites manquent (données nécessaires pour la sélection avec PPT) et 3) parcs dont la forme allongée présente des problèmes particuliers d'accès et réduit la signification des évaluations des expositions au bruit antérieures. Dans ces classes, le taux d'échantillonnage était

de beaucoup inférieur à celui utilisé dans les autres strates. C'est dans la strate urbaine que l'on observe le taux d'échantillonnage le plus faible (1 sur 79).

La strate à tirage complet comprenait les sept parcs qui, à cause des prescriptions de la loi, devaient être inclus dans l'étude. De plus, elle comprenait les parcs dont les taux de visites globaux (annuels) étaient suffisamment élevés pour assurer leur tirage dans l'échantillon obtenu au premier degré. La liste des 39 parcs dans l'échantillon est présentée dans le document 1; un 40$^{e}$ parc sélectionné (Grand Teton) a été éliminé de l'échantillon pour des raisons politiques.

Les calculs d'optimisation du plan ont mené, pour le premier degré, à une taille d'échantillon d'environ 40 parcs, donnant un total de 405 paires site-période (ou, dans le présent cas, de paires sortie-jour) choisies parmi les parcs très achalandés et peu achalandés. Comme on le décrit dans la sous-section 3.3, 15 paires sortie-jour ont été choisies dans chacun des trois parcs très achalandés et 10 dans chacun des 36 parcs peu achalandés dans l'échantillon. Pour une optimisation exacte, il faudrait utiliser les composantes de la variance pour la variance entre les parcs et pour la variance à l'intérieur des parcs. Ces composantes de la variance ont été obtenues de façon approximative à partir de données tirées d'une enquête antérieure et des données mensuelles sur les visites, qui sont disponibles pour toutes les zones administrées par le NPS.

### 3.2 Choix des blocs de deux mois

Bien que le choix des blocs de deux mois ait été effectué avec probabilités proportionnelles à la taille (PPT), on a aussi tenu compte d'exigences pratiques.

Premièrement, il était souhaitable que la formation des employés dans chacun des parcs s'effectue immédiatement avant la collecte des données. On a constitué des grappes géographiques de parcs afin que les moniteurs puissent visiter tous les parcs d'une grappe au cours d'un seul voyage. Plus précisément, les 39 parcs dans l'échantillon ont été groupés en 14 grappes de ce genre. Cette exigence a mené à la sélection d'une période de deux mois pour chaque grappe de parcs. Ainsi, quatorze blocs de deux mois ont été choisis, un pour chaque grappe. La mesure de taille utilisée pour chaque sélection était le nombre global de visites dans tous les parcs de la grappe.

Le document 1 montre les parcs échantillonnés dans chaque grappe ainsi que le bloc de deux mois choisi pour la grappe (c.-à-d. pour chaque parc dans la grappe). Trois strates, de groupes de grappes et par conséquent de parcs, ont été formées pour le choix des blocs de deux mois:

a) Dans la strate 1, "plein été" (comprenant 3 grappes), la base de sondage constituée de blocs de deux mois ne renfermait que la période de pointe de l'été, soit juillet et août, qui était alors choisie avec probabilités égales à un pour ces grappes;

b) Dans la strate 2, "début et fin de l'été" (composée de 7 grappes), la base de sondage constituée de blocs de deux mois renfermait 11 blocs de deux mois qui se chevauchaient;

c) Dans la strate 3, "printemps et automne" (comprenant 4 grappes), la base de sondage constituée de blocs de deux mois renfermait 5 blocs de deux mois qui ne se chevauchaient pas.

**Document 1**

Grappes de parcs, strates du deuxième degré et blocs de deux mois sélectionnés

| | | Période choisie |
|---|---|---|
| **Strate 1: Plein été*** | | |
| Grappe 2: | Mount Rainier, N. Cascades, Olympic | Juillet-août |
| Grappe 4: | Glacier, Yellowstone | Juillet-août |
| Grappe 10: | Sleeping Bear, Perry's Victory | Juillet-août |
| **Stratum 2: Début et fin de l'été** | | |
| Grappe 3: | Lassen, Yosemite, Kings Canyon/Sequoia | Août-septembre |
| Grappe 5: | Dinosaur, Rocky Mtn., Mt. Rushmore | Juin-juillet |
| Grappe 6: | Glen Canyon, Grand Canyon, Walnut Canyon | Juin-juillet |
| Grappe 8: | Bandelier, Lake Meredith | Mai-juin |
| Grappe 11: | Cape Cod, Delaware Gap, Gettysburg | Août-septembre |
| Grappe 12: | Shenandoah, Fredericksburg, Assateague | Juin-juillet |
| Grappe 13: | Great Smoky, Cape Hatteras, Fort Sumter | Juin-juillet |
| **Stratum 3: Printemps et automne** | | |
| Grappe 1: | Haleakala, Hawaii Volcanoes | Mars-avril |
| Grappe 7: | Lake Mead, Saguaro, Casa Grande | Mars-avril |
| Grappe 9: | Hot Springs, Wilson's Creek, Buffalo | Octobre-novembre |
| Grappe 14: | Cumberland Isd., Canaveral, Everglades, Gulf Island | Mars-avril |

*Choix de juillet-août avec probabilités égales à l'unité pour chaque grappe dans cette strate.

Deux raisons justifiaient cette stratification temporelle: elle permettait d'assurer que la collecte des données s'effectue pendant toute l'année et de distinguer les parcs où la grande majorité des visites ont lieu pendant l'été de ceux où la répartition des visites est plus uniforme.

Nous avons utilisé des procédures différentes, décrites ci-après, pour choisir les blocs de deux mois dans les deux dernières strates.

1) Pour les grappes de parcs dans la première de ces deux strates (début et fin de l'été), onze périodes qui se chevauchaient étaient incluses dans la base de sondage temporelle; janvier-février, février-mars, ..., novembre-décembre. Une de ces périodes était alors choisie avec probabilités proportionnelles à la taille (PPT). Il faut remarquer que, pour cette strate, chaque mois était inclus dans deux périodes faisant partie de la base de sondage sauf janvier et décembre. La probabilité de sélection de ces deux mois d'hiver était donc encore réduite (davantage que la probabilité déjà faible attribuée aux périodes hivernales avec la procédure PPT).

2) Pour les grappes de parcs dans la dernière de ces deux strates (printemps et automne), cinq périodes de deux mois qui ne se chevauchaient pas constituaient la base de sondage temporelle: janvier-février, mars-avril, mai-juin, septembre-octobre, novembre-décembre. Il faut remarquer que la période de deux mois de l'été, juillet-août (c.-à-d. le plein été), était exclue de la base de sondage pour ces parcs afin d'assurer le choix d'autres mois ne faisant pas partie de l'été. Pour chaque grappe dans cette strate, une parmi ces cinq périodes de deux mois était choisie avec PPT.

### 3.3 Troisième degré d'échantillonnage

Au troisième degré, les unités d'échantillonnage étaient les paires sortie-jour. Il est plus facile d'envisager l'échantillonnage au troisième degré comme la combinaison de deux choix indépendants: a) le choix de jours dans la période (bloc de deux mois) tirée au deuxième degré pour le parc et b) le choix des sorties à partir d'une liste de sorties précisées pour le parc. L'attribution aléatoire des jours échantillonnés aux sorties échantillonnées constituait le dernier degré.

L'échantillon des jours était stratifié par jours de semaine par opposition aux jours de fin de semaine (y compris les principaux jours fériés). Les jours échantillonnés étaient choisis avec probabilités égales dans chacune de ces deux strates. L'échantillon des sorties a été choisi avec probabilités proportionnelles à la taille (PPT). La mesure de taille attribuée à chaque sortie était l'utilisation relative de la sortie parmi toutes les sorties relevées pour le parc choisi. Cette mesure de l'utilisation a été obtenue avec l'aide d'employés de chacun des parcs.

Pour le plan d'échantillonnage du troisième degré, on faisait la distinction entre deux groupes de parcs désignés comme étant très achalandés (trois parcs: Grand Canyon et les deux parcs d'Hawaï) et peu achalandés (les 36 autres parcs de l'échantillon). La taille des échantillons pour les parcs peu achalandés était de 10 sorties et de 10 jours et, par conséquent, de 10 paires sortie-jour. Dans les trois parcs très achalandés, 15 sorties et 15 jours ont été choisis. Pour les parcs peu achalandés, nous avons utilisé une attribution égale entre la strate des jours de semaine et celle des jours de fin de semaine: 5 jours de semaine et 5 jours de fin de semaine ont été choisis de façon indépendante dans chacun de ces (36) parcs dans l'échantillon. Pour les parcs très achalandés, l'attribution était à peu près égale puisqu'on a choisi 7 jours de fin de semaine et 8 jours de semaine.

### 3.4 Quatrième degré d'échantillonnage

Au cours du quatrième degré, les visiteurs des parcs ont été interviewés sur le vif aux sorties choisies lors des jours sélectionnés. Un échantillon aléatoire systématique de visiteurs (ou de groupes de visiteurs) quittant le parc a été choisi avec un pas de sondage fixe pour chaque unité (paire sortie-jour) sélectionnée au troisième degré. On a laissé le pas de sondage varier d'un jour à l'autre afin de tirer profit de l'expérience des jours précédents et de la variabilité dans le nombre de visiteurs entre les jours et les sorties. Chaque visiteur faisant partie des groupes choisis admissibles était soumis à un filtrage afin de s'assurer qu'il était admissible, puis interviewé s'il l'était (visiteur adulte).

### 3.5 Enquête sur l'utilisation des installations de l'arrière-pays et enquête postale

Le même échantillon de parcs (UPÉ) a été utilisé pour une enquête postale auprès des utilisateurs des installations de l'avant-pays et de celles de l'arrière-pays pendant la période de deux mois sélectionnée pour le parc. L'échantillon de l'enquête sur l'utilisation des installations de l'arrière-pays était limité au sous-ensemble des parcs échantillonnés pour lesquels il y avait une certaine utilisation de l'arrière-pays. Dans chacun de ces parcs échantillonnés, la base de sondage (du troisième degré) pour cette composante de l'enquête était fondée sur les permis de circulation dans l'arrière-pays délivrés pendant la période de la collecte des données (bloc de deux mois) établie pour le parc. Le dernier échantillon d'utilisateurs d'installations de l'arrière-pays était alors choisi avec probabilités égales à partir des listes de permis fournies par le personnel du parc. Un sous-échantillon de visiteurs choisis lors des interviews sur le vif était aussi sélectionné pour l'enquête postale.

L'échantillon utilisé pour l'enquête postale était fondé sur le même échantillon, pour le troisième degré, de paires sortie-jour choisies pour les interviews sur le vif. Pour chaque paire sortie-jour choisie, on a obtenu un sous-échantillon composé d'un nombre fixe de groupes parmi les groupes répondant aux interviews sur le vif (ce nombre était de 15 dans les parcs très achalandés et de 10 dans les parcs peu achalandés). Un autre degré de sous-échantillonnage fût de choisir une personne parmi les répondants dans chaque groupe sous-échantillonné. Ce sous-échantillonnage des groupes et des personnes s'effectuait avec échantillonnage aléatoire stratifié pour contrôler la composition démographique de l'échantillon final.

## 4. PLAN D'ÉCHANTILLONNAGE DE L'ENQUÊTE PORTANT SUR LES TROIS RIVIÈRES

### 4.1 Construction de la base de sondage et stratification

Tout d'abord, RTI et Terrestial Environment Specialists (TES) ont construit une base de sondage fondée sur un relevé de tous les sites dans la région visée par l'enquête portant sur les trois rivières. Puis un échantillon stratifié à plusieurs degrés a été choisi indépendamment pour chacun des douze mois de l'étude. Les unités d'échantillonnage du premier degré étaient les paires site-période et celles du deuxième degré, les personnes s'adonnant à des activités récréatives dans les paires site-période choisies. Au premier degré de l'échantillonnage des périodes et des sites, on a eu recours à la stratification temporelle et à la stratification dans l'espace.

La stratification primaire dans la dimension temporelle a été effectuée selon le mois, et la stratification primaire des sites s'est faite selon le type d'utilisation: les sites (points d'accès) ont été classés comme sites de navigation

de plaisance, sites de pêche sportive ou parcs. Chaque strate primaire a été divisée en six régions géographiques, ou groupes, définies comme étant des segments de rivière entre des écluses. La strate de la pêche sportive a ensuite fait l'objet d'une sous-stratification en fonction de l'intensité d'utilisation prévue, soit intensité faible ou intensité élevée: les sites à intensité élevée étant ceux en aval des écluses et des barrages.

Chaque échantillon mensuel de paires site-période a été choisi par échantillonnage aléatoire stratifié. Les avantages de la sélection d'échantillons mensuels indépendants sont: la possibilité de calculer des estimations mensuelles et saisonnières et la possibilité de modifier certaines caractéristiques du plan d'un mois à l'autre.

Par exemple, ce plan a permis de modifier la base de sondage spatiale d'un mois à l'autre par l'ajout ou la suppression de sites. En particulier, la strate de la navigation de plaisance pour les mois d'hiver (de novembre à avril) a été limitée aux rampes de mise à l'eau qui étaient ouvertes pendant ces mois. De plus, plusieurs sites de pêche sportive inclus dans la base de sondage pour les premiers mois se sont révélés inaccessibles et ont été supprimés de la base de sondage pour les échantillons des mois ultérieurs.

Parmi les autres caractéristiques du plan qui ont changé au cours des mois, mentionnons: les taux d'échantillonnage, au deuxième degré, pour la sélection des utilisateurs admissibles ainsi que les périodes de collecte des données le matin, l'après-midi et en soirée pour les sites de chaque type. Des périodes de collecte des données variables ont été utilisées au cours de différents mois de l'étude. Ces périodes étaient définies à l'aide de renseignements sur le lever et le coucher du soleil ainsi que sur les habitudes d'utilisation prévues pour les divers types de sites. Les mois d'hiver (de novembre à avril) incluaient deux périodes par jour, alors que les mois d'été (de mai à octobre) en comprenaient trois.

La (sous) stratification temporelle de chaque échantillon mensuel a été effectuée selon les jours de fin de semaine par opposition aux jours de semaine. Il vaut la peine de faire remarquer que la strate des jours de fin de semaine incluait aussi les principaux jours fériés.

### 4.2 Sélection de l'échantillon

Comme on l'a signalé plus haut, nous avons choisi, au premier degré, un échantillon indépendant de paires site-période pour chacun des 12 mois de l'étude. Chaque échantillon mensuel comprenait deux composantes: a) un échantillon aléatoire stratifié de "*n*" sites et b) un échantillon aléatoire stratifié de "*n*" périodes.

Après la sélection des périodes pour chaque mois, les sites échantillonnés ont été attribués aléatoirement aux périodes sélectionnées. L'attribution des sites aux périodes a été faite entièrement au hasard pour les mois de février à juin, mais elle a été modifiée au cours des mois ultérieurs. À partir de juillet, un échantillon des périodes était choisi indépendamment pour chaque strate de type de site, l'attribution aléatoire se faisant dans la strate. La répartition des périodes échantillonnées (p. ex., le nombre de périodes en matinée et le nombre de périodes en soirée incluses dans l'échantillon) variait d'une strate à l'autre. Avec cette méthode plus flexible, on a pu, par exemple, attribuer un nombre relativement plus élevé de sites de pêche sportive aux périodes du matin et plus de sites de navigation de plaisance aux périodes de l'après-midi. Le document 2 présente la taille des échantillons, sites et périodes, utilisés dans l'attribution aléatoire chaque mois.

**Document 2**

Taille des échantillons utilisés pour l'attribution aléatoire des sites aux périodes pour chaque échantillon mensuel de l'enquête portant sur les trois rivières

| Mois | Globale* | Taille de l'échantillon | | Parcs | Marinas | Régates | |
|---|---|---|---|---|---|---|---|
| | | Navigation de plaisance | Pêche sportive | | | Navigation de plaisance | Marinas |
| 1 | | 8 | 12 | 4 | | | |
| 2 | 20 | | | | | | |
| 3 | 28 | | | | | | |
| 4 | 28 | | | | | | |
| 5 | 36 | | | | | | |
| 6 | 36 | | | | | | |
| 7 | | 17 | 22 | 6 | 12 | | |
| 8 | | 10 | 22 | 6 | 6 | 6+ | 6+ |
| 9 | | 10 | 14 | 6 | 6 | | |
| 10 | | 10 | 12 | 4 | | | |
| 11 | | 8 | 12 | 4 | | | |
| 12 | | 8 | 12 | 4 | | | |

* Pour ces cinq mois, l'attribution a eu lieu pour l'ensemble des sites et des périodes échantillonnés (il N'y a PAS eu de regroupement selon le type de site). Pour les autres mois, l'attribution a été faite dans la strate de chaque type, processus qui comprenait le choix d'échantillons indépendants de périodes pour chaque type.

\+ Pour août, on a tout d'abord attribué séparément les sites des régates aux périodes. Puis, les paires site-période échantillonnées associées aux régates ont été déplacées soit vers la strate de la navigation de plaisance, soit vers celle des marinas, selon le type de site.

Les taux d'échantillonnage, du deuxième degré, étaient précisés pour chacune des trois strates primaires avant la collecte des données pour chaque mois. Ces taux propres à la strate étaient déterminés d'après l'expérience acquise au cours des mois antérieurs, et ils étaient fournis aux intervieweurs sur le terrain avec le calendrier de collecte des données du mois. Les personnes ont été choisies par échantillonnage aléatoire systématique dans chaque paire site-période.

### 4.3 Enquête sur les marinas et les événements spéciaux

Une enquête sur les marinas a été réalisée de juin à septembre. On a utilisé une base de sondage composée de 48 marinas fondée sur le relevé effectué par TES, qui a été mis à jour à la fin de mai 1992.

La méthode d'échantillonnage utilisée pour le mois d'introduction de l'enquête sur les marinas, juin différait légèrement de celle qui a été utilisée pour les échantillons des mois suivants. L'échantillon de juin était un supplément de 12 marinas combiné à un échantillon de 12 jours. Les échantillons de marinas pour les mois de juillet à septembre ont été choisis en utilisant la base de sondage

des marinas comme une quatrième strate. Les procédures de sélection étaient alors semblables à celles utilisées dans les trois autres types de strate; plus précisément, a) un échantillon de "$n$" marinas a été choisi, b) un échantillon de "$n$" périodes a été choisi et c) les sites échantillonnés ont été attribués aléatoirement aux périodes choisies.

Il faut mentionner que certaines des marinas étaient aussi incluses dans la strate de la navigation de plaisance. Dans de tels cas, deux unités distinctes de la base de sondage ont été créées pour le même site. Cette situation s'est aussi produite pour certains sites utilisés à la fois pour la navigation de plaisance et pour la pêche sportive, et les sites de ce genre ont été inclus dans les deux strates.

En plus de l'enquête sur les marinas, nous avons relevé des événements spéciaux qui ont eu lieu dans la région et au cours de la période de douze mois visée par l'enquête. La plupart de ces événements ont été traités d'une façon qui ressemble à celle utilisée pour les jours de fin de semaine et les jours fériés en les attribuant à une strate qui sera suréchantillonnée. Une catégorie qui présentait un intérêt spécial était celle des régates, qui ont lieu pendant les mois d'été. Pour deux échantillons mensuels (juillet et août), nous avons relevé les dates des régates ainsi que les sites où chacune d'entre elles allaient se dérouler. Nous avons alors construit une (cinquième) strate distincte pour inclure ces paires site-période. Au cours du premier degré de l'échantillonnage, la répartition de l'échantillon dans la strate des régates reflétait le suréchantillonnage souhaité pour cette strate. Comme on le voit dans le document 2, la procédure d'échantillonnage utilisée pour choisir les paires site-période (unités du premier degré) à partir de la strate des régates différait aussi de celle employée dans les quatre autres strates. Des paires site-période ont été choisies directement en une étape depuis le sous-ensemble des paires site-période dans la strate, c.-à-d. qu'aucune attribution aléatoire n'a été requise.

## 5. PONDÉRATION POUR LES ENQUÊTES

### 5.1 Pondération pour l'enquête du NPS

Les poids d'échantillonnage ont tout d'abord été calculés pour chacune des trois premiers degrés de sélection. Pour le premier degré, le poids d'échantillonnage pour chaque parc dans l'échantillon était l'inverse de la probabilité de sélection du parc. Pour le deuxième degré, le poids d'échantillonnage pour chaque bloc de deux mois dans l'échantillon était calculé de la même façon. La somme globale (ou dans une strate) des poids d'échantillonnage pour le premier degré était le nombre de parcs dans la base de sondage (ou dans la strate).

Pour le troisième degré, les poids pour les paires sortie-jour dans l'échantillon étaient le produit de deux facteurs associés au choix des jours et des sorties. Il faut remarquer que pour chaque parc et pour chaque bloc de deux mois sélectionnés, la somme du premier ensemble de poids dans une strate temporelle (jours de fin de semaine par opposition à jours de semaine) est le nombre de jours dans la strate, et la somme du dernier ensemble de poids est le nombre de sorties relevées pour le parc. Ces poids ont été rajustés afin de tenir compte de la non-réponse qui s'est produite au troisième degré parce que, dans quelques parcs, la collecte des données n'a pas eu lieu pour certaines des paires sortie-jour sélectionnées. Dans tout parc où l'on a observé ce défaut au niveau de la collecte des données, on a rendu la somme des poids rajustés, pour le troisième degré, pour les paires sortie-jour actives égale à la somme des poids d'échantillonnage pour toutes les paires sortie-jour sélectionnées dans le parc.

Pour le quatrième degré, les poids ont été calculés au niveau d'un groupe et au niveau d'une personne. Les poids au niveau d'un groupe sont attribués à tous les groupes participants de paires sortie-jour dans l'échantillon et ils ont la même valeur pour les groupes dans la même paire sortie-jour. De même, les poids au niveau de la personne ont été attribués à toutes les personnes interviewées sur le vif dans une paire sortie-jour dans l'échantillon. Les poids d'échantillonnage, pour le quatrième degré, ont été calculés comme l'inverse du taux de sondage précisé pour la paire sortie-jour dans l'échantillon. Ces poids ont alors été rajustés afin de tenir compte de la non-réponse au niveau du groupe et de la personne.

L'échantillon de l'enquête postale était basé sur l'échantillon, pour le troisième degré, de paires sortie-jour choisies pour les interviews sur le vif. Pour chaque paire sortie-jour sélectionnée dans un parc peu achalandé, on a tout d'abord choisi, avec probabilités égales, dix des groupes participants; puis, un sous-échantillon d'une personne a été choisi, parmi tous les répondants aux interviews sur le vif, dans chaque groupe sélectionné. Une procédure semblable a été utilisée dans les parcs très achalandés, sauf que, dans ce cas, on a choisi 15 groupes plutôt que 10 par paire sortie-jour.

Le poids d'échantillonnage pour chaque enregistrement de l'enquête postale est le produit de WTB = nombre de répondants aux interviews sur le vif dans le groupe et de WTA = nombre de groupes participants/10 [pour les parcs peu achalandés]. Dans le cas des parcs très achalandés, le dénominateur de WTA est 15 plutôt que 10.

Ces poids ont été rajustés, pour tenir compte de la non-réponse lors de l'enquête postale, à l'aide de paires sortie-jour (dans un parc) utilisées comme classes de pondération. Ainsi, la somme des poids rajustés pour tous les répondants provenant de la même paire sortie-jour est identique à la somme des poids (non rajustés) pour toutes les personnes ayant répondu aux interviews sur le vif dans la paire sortie-jour.

Pour l'enquête portant sur l'utilisation des installations de l'arrière-pays, la base de sondage pour chaque parc admissible (dans le sous-ensemble des parcs échantillonnés avec utilisation d'installations de l'arrière-pays) était fondée sur les listes de permis délivrés pendant la période de collecte des données: les permis admissibles étaient associés à des dates de sortie pendant cette période. Un échantillon de cinq groupes par jour a été choisi dans chaque parc à partir de l'ensemble des permis liés à ce jour.

Un sous-échantillon d'une personne a été choisi dans chaque groupe. Le calcul des poids équivaut à celui qui est fait pour l'enquête postale avec

$$\text{WTBACK} = \text{BACKA}*\text{BACKB},$$

où, pour chaque jour:

BACKA = (nombre de groupes liés au jour)/5
BACKB = nombre de personnes dans le groupe.

Les poids d'analyse pour l'enquête portant sur l'utilisation des installations de l'arrière-pays découlaient des rajustements pour la non-réponse apportés en utilisant les jours comme classes de pondération.

### 5.2 Pondération pour l'enquête portant sur les trois rivières

Au moment de la sélection de chaque échantillon mensuel de paires site-période, nous avons calculé des poids d'échantillonnage qui reflétaient la probabilité de sélection des paires site-période. Les poids initiaux étaient le produit de deux ensembles de poids calculés pour chaque échantillon mensuel: a) les poids attribués à chaque site dans l'échantillon aléatoire stratifié de sites du type donné (navigation de plaisance, pêche sportive, marina et fréquentation des parcs) et b) les poids attribués à chaque période dans l'échantillon aléatoire stratifié des périodes. On a alors extrapolé ces poids pour tenir compte de l'attribution aléatoire des sites échantillonnés aux périodes (ou inversement). Ainsi, pour chaque mois, la somme des poids pour les paires site-période était égale au nombre de combinaisons de paires site-période dans la base de sondage.

Le processus de rajustement des poids a commencé avec les poids d'échantillonnage associés aux paires site-période sélectionnées. Un rajustement initial a été apporté aux poids, pour le premier degré, afin de tenir compte des paires site-période qui avaient été trouvées inadmissibles ou pour lesquelles des questionnaires manquaient (p. ex., questionnaires qui n'avaient pas été expédiés par les employés travaillant sur le terrain). Pour ce rajustement, au premier degré, nous avons utilisé les strates type par mois comme classes de pondération. C'est-à-dire que nous avons rendu la somme des poids rajustés sur l'ensemble réduit égale à la somme des poids non rajustés sur tout l'échantillon dans la classe type par mois. Un rajustement final a été apporté au niveau du répondant pour refléter a) le pas de sondage systématique utilisé dans les paires site-période et b) la non-réponse à l'enquête au niveau des particuliers.

Dans le cadre des procédures de vérification de la pondération, nous avons calculé la somme des poids d'analyse finals pour tout le fichier et aussi par mois et par type de site. La somme des poids devrait correspondre, approximativement, au nombre total estimé de personnes utilisant des installations de loisirs qui quittent les sites relevés pendant la période de collecte des données pour chaque mois et pour chaque type de site.

## 6. CONCLUSION

Dans cet article, nous avons décrit le plan d'échantillonnage utilisé pour deux enquêtes, menées auprès de personnes qui utilisent des installations de loisirs et qui partagent un certain nombre de caractéristiques. Les plans d'échantillonnage comprennent l'échantillonnage dans le temps ainsi que dans l'espace; des paires site-période sont choisies au degré qui précède l'échantillonnage des visiteurs. Nous construisons alors une base de sondage spatiale en même temps qu'une base de sondage temporelle. Une fois que les sites sont choisis à partir de la première base de sondage et les périodes, à partir de la dernière, les sites échantillonnés sont attribués aléatoirement aux périodes échantillonnées (ou inversement). Les poids d'échantillonnage doivent tenir compte de cette étape additionnelle de randomisation. Les conclusions de l'enquête auprès des visiteurs effectuée pour le NPS sont décrites dans le rapport final de l'enquête (National Park Service 1994). Cette analyse incluait divers modèles de régression employés pour évaluer l'incidence liée au fait d'entendre et de voir des avions survoler les zones administrées par le NPS.

Du point de vue temporel, les deux enquêtes portent sur différentes périodes correspondant à toute l'année, c'est-à-dire qu'un utilisateur aura une probabilité positive d'être choisi pour toute période de l'année. Les deux études comprennent aussi une stratification temporelle pour refléter les habitudes d'utilisation et pour augmenter l'efficacité de l'échantillonnage.

Du point de vue spatial, alors que l'échantillon utilisé dans le cadre de l'enquête menée pour le National Park Service était un échantillon national de visiteurs des zones administrées par le NPS, l'échantillon employé pour l'enquête portant sur les trois rivières englobait une région plus limitée. Toutefois, dans les deux enquêtes, on faisait une distinction entre différents types d'utilisateurs. Pour l'enquête du NPS, les utilisateurs des installations de l'arrière-pays et ceux des installations de l'avant-pays étaient choisis (au troisième degré) dans deux strates distinctes, les détenteurs de permis étant visés par l'enquête portant sur l'utilisation des installations de l'arrière-pays. Dans le cas de l'enquête portant sur les trois rivières, les visiteurs étaient classés selon qu'ils s'intéressaient surtout à la pêche sportive ou à la navigation de plaisance ou qu'ils fréquentaient davantage les parcs, et les sites étaient stratifiés de la même façon. Ce plan d'échantillonnage a permis le calcul d'estimations précises selon le type et la saison.

La dernière unité d'échantillonnage pour l'enquête sur les personnes qui utilisent des installations de loisirs est la visite; ainsi, les visiteurs peuvent avoir plusieurs chances d'être inclus dans l'échantillon dans la mesure où ils utilisent plusieurs fois les zones visées par l'enquête. Il vaut la peine de souligner que cette structure est compatible avec les objectifs d'enquêtes de ce genre.

Les poids d'échantillonnage tenaient compte de la sélection des unités temporelles et spatiales à chaque degré ainsi que de l'opération d'attribution aléatoire. Les échantillons étaient conçus de façon à minimiser les effets de la pondération inégale sur les variances de l'enquête. On a tenu

compte de l'importance possible de ces effets lorsqu'on a combiné différentes composantes de l'enquête. Voici des exemples pour les deux enquêtes, soit la combinaison:

a) des composantes arrière-pays et avant-pays de l'enquête postale du NPS et
b) des personnes qui pratiquent la pêche sportive, la navigation de plaisance et qui fréquentent des parcs dans l'enquête portant sur les trois rivières.

Il faudrait aussi signaler certains inconvénients de ce type de plan d'échantillonnage. Alors que l'échantillonnage dans le temps et l'attribution aléatoire introduisent un élément de rigueur statistique et étendent la gamme d'inférences statistiques valides, la méthodologie peut être perturbée si les intervieweurs sur le terrain changent la date qui correspond à un emplacement échantillonné parce que l'accès à cet emplacement peut être difficile pendant la période précisée ou pour d'autres raisons. À titre d'exemples notables de situations de ce genre que l'on a rencontrées au cours de l'enquête du NPS, mentionnons des ouragans, des parcs fermés à cause de la présence de fugitifs recherchés par la justice, des lancements de navettes spatiales et de graves tempêtes de neige. Alors qu'on peut minimiser certaines de ces situations au moyen de la stratification temporelle et de la répartition de l'échantillon, le statisticien ne peut manifestement rien faire pour d'autres. Le statisticien qui s'occupe d'échantillonnage devrait participer à la formation des intervieweurs afin d'insister sur le fait qu'il faut éviter, à tout prix, de modifier le plan d'échantillonnage et il devrait surveiller toutes les modifications qui sont apportées.

## REMERCIEMENTS

Les auteurs remercient le rédacteur associé ainsi que les arbitres pour leurs commentaires, qui ont aidé à améliorer la présentation de l'article.

## BIBLIOGRAPHIE

GOUGH, J.H., et GHANGURDE, P.D. (1977). Survey of Canadian residents returning by land. *Techniques d'enquête*, 3, 215-231.

KALTON, G. (1991). L'échantillonnage des flux de populations humaines mobiles. *Techniques d'enquête*, 17, 197-210.

KISH, L., LOVEJOY, W., et RACKOW, P. (1961). A Multi-stage probability sample for traffic surveys. *Proceedings of the Social Statistics Section, American Statistical Association*, 227-230.

LEVY, M.R. (1983). The methodology and performance of election day polls. *Public Opinion Quarterly*, 54-67.

NATIONAL PARK SERVICE (1994). Survey of Visitors to National Park Service Areas: Survey Findings. Report 94-2, No. 290940.12.

SUDMAN, S. (1980). Improving the quality of shopping center sampling. *Journal of Marketing Research*, 17, 423-431.