

SYNTHESE DU FOCUS SUR LA METHODOLOGIE DES FUSIONS UTILISEES DANS LE CADRE DU DISPOSITIF INTERNET GLOBAL

A la demande du marché, Médiamétrie a confié au CESP une série de focus ciblés sur son nouveau dispositif Internet Global mis en place pour produire les résultats d'audience internet sur trois écrans (ordinateur, smartphone et tablette). Dans le cadre de cette mission, le CESP a mené un premier Focus sur le panel Tablette et la mesure d'audience d'internet via une tablette publié en mai 2016, deux Focus sur la mesure Internet Mobile, en particulier sur le recrutement du nouveau panel Mobile en décembre 2016 et sur l'« hybridation » des résultats en avril 2017. Le présent rapport porte sur les analyses et les conclusions du Comité Scientifique sur la méthodologie des fusions entre les trois panels Internet de Médiamétrie. Un cinquième Focus sur les dispositifs de mesure du panel Mobile et du panel Tablette (recueil des données de navigation via un *meter*, une application, un serveur *proxy*, etc.) est en cours.

Afin de communiquer un avis sur la nouvelle méthodologie des fusions dans la foulée de la publication des premiers résultats d'audience, le CESP et Médiamétrie se sont accordés pour limiter le champ du contrôle et les premières analyses du CESP. Sont couverts par ce Focus :

- le recrutement du dispositif WEBALL (ensemble des individus constituant les trois panels Internet)
- le suivi et la gestion des trois panels (relances, règles d'exclusion...)
- l'analyse de la qualité des échantillons de chacun des trois panels (taille, structure de l'échantillon, taux de *churn*...)
- l'examen de la méthodologie générale retenue par Médiamétrie pour fusionner les trois panels
- le contrôle de la fusion « Mobilité » -i.e. le rapprochement entre le panel Mobile et le panel Tablette- sur trois strates : femmes 35-49 ans, hommes 50 ans ou plus inactifs et femmes 50 ou plus actives.

Les étapes de pré-traitements conduites par Médiamétrie au niveau des panels Internet¹ n'ont pas été contrôlées dans ce Focus et pourront faire l'objet de Focus complémentaires. En particulier, le Comité Scientifique souligne l'importance de conduire un contrôle sur l'imputation de données https mise en œuvre par Médiamétrie depuis juin 2017 pour pallier les problèmes rencontrés par le *meter* NetSight dans la collecte du surf sécurisé de panélistes qui utilisent Chrome comme navigateur (plus de 50% des internautes utilisent ce navigateur sur leur ordinateur²).

Le Comité Scientifique prend acte des difficultés inhérentes à la complexité technique et aux évolutions technologiques permanentes de l'univers à étudier. Il observe notamment que plusieurs types de terminaux ne peuvent pas être pris en compte dans la mesure actuelle (exemples : ordinateurs avec un système d'exploitation autre que MAC ou Windows, les tablettes et smartphones sous Windows, smartphones sous Android versions 3 et antérieures, iPhones de versions iOS 5 et antérieures, mesure partielle des apps. sous iPhone iOS 8 ou plus³), ce qui introduit un biais dans la mesure des usages d'internet.

Au préalable, le Comité Scientifique souligne qu'il n'a pas été en mesure de se prononcer sur certains points méthodologiques et leur mise en œuvre faute d'informations communiquées par Médiamétrie. Il regrette le manque de fluidité dans la coopération entre les équipes et le non-respect des délais pour fournir les éléments demandés. Le Comité Scientifique souligne aussi l'importance de la fourniture au CESP d'une documentation complète et du code des programmes employés, d'autant plus que la méthodologie auditée est nouvelle et complexe.

¹ A titre d'illustration, les pré-traitements faits au panel Ordinateur tels que l'exclusion temporaire de panélistes « atypiques », le traitement des « *outliers* » (i.e. l'élimination de visites par sous-domaine), la fusion « Domicile et Lieu de travail » ne sont pas couverts par le présent Focus. Le contrôle des redressements « hybride » est également hors du champ du Focus.

² 52% en février 2018 d'après la source StatCounter.

³ A noter, depuis janvier 2017, Médiamétrie a recours à un algorithme d'inférence développé par RealityMine pour identifier les applications étant en focus dans le cas d'un smartphone iOS 8 ou plus. Cet algorithme n'est pas documenté.

A partir de ses différents contrôles et analyses conduits sur l'année 2017 et en particulier sur les données utilisées par Médiamétrie pour déterminer le choix de sa méthodologie de fusions, le CESP fait les observations suivantes :

Concernant la méthodologie mise en œuvre par Médiamétrie pour rapprocher les trois panels Internet

- Le choix d'une approche par fusions avec transfert de poids, consistant à conserver les poids initiaux des panélistes, nécessite une maîtrise de la dispersion des poids de redressement de chacun des panels Internet mis en entrée des fusions. Cela passe par la maîtrise du recrutement et de la gestion des trois panels Internet afin de garantir la qualité des échantillons, en particulier celle du sous-échantillon des individus qui sont mesurés sur plusieurs types de terminaux.
- Le « panel Trois Ecrans » à partir duquel les résultats d'audience sont désormais calculés est issu de fusions⁴ et correspond à un échantillon d'individus virtuels. La construction de ce « panel » s'appuie principalement sur le sous-échantillon des panélistes bi- et tri-mesurés, 2 967 panélistes répondants actifs en juillet 2017, qui sert de socle à l'élaboration des modèles de passage entre panels (les transformations procustéennes).

Le CESP considère satisfaisant :

- Le principe d'une méthode de rapprochement par transfert de poids qui présente l'avantage de fusionner plusieurs études tout en conservant les objectifs de chacune des études prises séparément. Cette méthode permet notamment de conserver les structures par panel Internet et de maintenir les contraintes de redressement propres à chaque panel. Toutefois, le Comité Scientifique souligne que les éléments communiqués n'ont pas permis de rendre compte de la démarche scientifique de Médiamétrie. Notamment, les choix techniques effectués à chaque étape sont insuffisamment étayés et testés pour évaluer leurs impacts sur les résultats.

Le CESP émet des réserves sur :

- La faiblesse des effectifs de panélistes bi- ou tri- mesurés pour la construction des modèles utilisés pour fusionner les panels Internet (entre 41 et 201 panélistes par strate en octobre 2017).

Le CESP recommande :

- D'augmenter la taille du sous-échantillon des panélistes bi- ou tri- mesurés pour améliorer les modèles de construction des fusions (transformations procustéennes) et de conduire des analyses complémentaires sur chaque strate afin d'apprécier l'hétérogénéité des profils et évaluer la robustesse de chaque modèle.
- D'étendre l'analyse conduite sur les trois strates sélectionnées par le CESP à l'ensemble des strates, et réaliser une validation croisée sur chacune des fusions pour vérifier la stabilité et la robustesse de la méthodologie.
- De mettre à disposition du CESP les éléments nécessaires pour estimer la précision des résultats d'audience (à partir d'une approche par *bootstrap*).
- De justifier par des résultats de tests le choix du nombre d'axes retenus pour l'analyse en composantes principales dans chaque strate.
- De tester la possibilité de ne pas considérer dans chaque strate le Top 100 global des brands, mais le Top 100 de chaque strate.

Le Comité Scientifique devra disposer de ces analyses pour émettre un avis sur les résultats produits par la nouvelle méthodologie Internet Global.

- De mener une analyse de la stabilité des résultats dans le temps.

⁴ 7 fusions entre le panel Mobile et le panel Tablette qui aboutissent à un premier panel virtuel, le « panel Mobilité ». 47 fusions entre le « panel Mobilité » et le panel Ordinateur.

Concernant le recrutement et la gestion du dispositif conçu par Médiamétrie pour recruter les panélistes Internet

Le CESP considère satisfaisants :

- Le principe d'une enquête de calage unique pour produire les structures de redressement de l'ensemble des populations Internet prises en compte dans l'univers du dispositif Internet Global.
- L'augmentation de la part des recrutements de panélistes mobinautes et de panélistes bi- ou tri- mesurés sur plusieurs types de terminaux.
- L'évolution à la baisse des taux de *churn* par panel en 2017.

Le CESP émet des réserves sur :

- L'utilisation de l'enquête Observatoire des Usages Internet - Téléphonie et Services Mobiles (OUI-TSM) pour le calcul des structures théoriques de l'univers mesuré compte tenu de la faiblesse des effectifs sur lesquels ces structures reposent (inférieurs à ceux des panels Internet).
- Les règles de recrutement des trois panels qui ont été définies sans quotas.
- Certaines règles d'exclusion de panélistes qui sont relatives au comportement que l'on souhaite mesurer dans le dispositif Internet Global, telle que l'exclusion temporaire de panélistes Ordinateur sur la base de leur activité⁵ ou de leur navigation⁶ sur internet. Le Comité Scientifique souligne que ces règles peuvent avoir un fort impact sur la taille et la représentativité des échantillons de panélistes contribuant aux calculs des audiences. En octobre 2017, 13% de l'échantillon du panel Ordinateur ont ainsi été exclus.
- L'absence d'homogénéité entre l'univers mesuré et l'univers théorique à cause de l'exclusion de certains terminaux qui restreignent l'échantillon étudié (ex : ordinateurs avec un système d'exploitation autre que Windows ou MAC, smartphones autres que iOS ou Android, tablettes Surface de Windows...).
- Concernant le panel Ordinateur, la représentativité du sous-échantillon WORK compte tenu des contraintes d'installation du *meter* sur des ordinateurs professionnels.

Le CESP ne peut se prononcer sur :

- La qualité des sources d'adresses utilisées par Médiamétrie pour le recrutement, sur lesquelles un focus est en cours.

Le CESP recommande :

- De préciser l'univers mesuré, notamment l'exclusion des tablettes sous Windows et des smartphones sous Windows, et l'impossibilité de discerner les applications en focus sous iOS 8 ou plus⁷.
- De conduire une étude sur les comportements et les usages Internet mobile des enfants entre 2 et 10 ans pour redéfinir l'âge d'entrée des individus dans l'univers mesuré.

⁵ Les panélistes Ordinateur répondants qui ont déclaré se connecter tous les jours ou presque depuis un ordinateur dans leur page de Profil et sont inactifs le mois en cours et le mois précédent le mois de mesure, sont temporairement exclus du mois de mesure et ne contribuent donc pas au calcul des audiences.

⁶ La règle d'exclusion des « atypiques joueurs » consiste à exclure de la base des répondants les panélistes Ordinateur qui ont visité un trop grand nombre d'entités marché différentes appartenant aux deux sous-catégories « Jeux Concours / Loteries » et « Coupons de réduction ».

⁷ A noter, depuis janvier 2017, Médiamétrie a recours à un algorithme d'inférence développé par RealityMine pour identifier les applications étant en focus dans le cas d'un smartphone iOS 8 ou plus. Cet algorithme n'est pas documenté.

- De revoir la gestion et le suivi des trois panels Internet :
 - mieux sélectionner les sources de recrutement pour maîtriser les biais de chacune d'elles et ne plus avoir recours à des règles d'exclusion de panélistes⁸.
 - définir une période probatoire avant d'intégrer les panélistes répondants dans l'échantillon pour la production des résultats d'audience. Cela devrait limiter la volatilité observée parmi les répondants d'une vague sur l'autre.
- D'adapter la procédure de relances afin de contacter les panélistes qui ne participent pas correctement à la mesure au sein des foyers des panels Ordinateur ou Tablette et ne plus avoir recours à des règles d'exclusion sur les incohérences entre des habitudes déclarées par un tiers dans la page de profil et un défaut d'usage observé.
- De rester vigilant sur le taux de *churn* de chacun des panels.
- De modifier le guide mis à disposition des utilisateurs dans la nouvelle interface de restitution des résultats, en précisant que les seuils sont calculés sur des individus virtuels et non des panélistes bruts.
- De fournir les effectifs bruts (nombre de panélistes) des 55 cibles standards et le seuil considéré comme minimum en termes d'effectif brut pour créer une cible dans l'interface.

Le CESP recommande aux utilisateurs :

- De ne pas comparer les résultats actuels avec des résultats datant d'avant le mois d'octobre 2017, qui correspond au mois du changement de méthodologie.
- D'être vigilants sur la significativité des résultats, en particulier compte tenu de l'extrême faiblesse des seuils de publication des résultats à la case (à partir de 50 individus virtuels, ce qui correspond en moyenne à environ 16 panélistes réels en effectif brut). A cet égard, le Comité Scientifique considère que l'utilisation du terme « panéliste » dans l'interface est impropre.
- De réfléchir à la pertinence de la définition d'internaute, basée sur une période de référence des 30 derniers jours pour un média essentiellement d'usage quotidien.

Le Comité Scientifique rappelle que les audits ou focus du CESP sont conduits à la demande de ses membres en vue d'accompagner la réalisation des mesures d'audience, d'assurer leur transparence, et d'œuvrer avec les instituts à leur amélioration et à leur adaptation aux évolutions du marché. Il est primordial que les conclusions du CESP soient communiquées au marché de manière synchrone avec la publication des résultats des études contrôlées. En particulier, dans le cas d'un nouveau dispositif de mesure d'audience, les conséquences des modifications et leur évaluation doivent être connues du marché en amont de l'utilisation des résultats.

Le Comité Scientifique,
le 28 mars 2018



⁸ Le CESP suggère d'élaborer un modèle prédictif de *churn* pour identifier en amont les panélistes qui auraient l'intention de quitter le panel afin de mettre en place des actions préventives au moment approprié.