

SYNTHESE DE L'AUDIT DU CESP SUR LE DISPOSITIF INTERNET GLOBAL EN 2018

A la demande du marché, le CESP audite le dispositif Internet Global de Médiamétrie pour produire les résultats d'audience internet sur trois écrans (ordinateur, smartphone et tablette). Ce dispositif mis en place à partir des données d'octobre 2017 s'appuie sur les trois panels Internet, qui produisaient précédemment les résultats d'audience par écran, et sur une nouvelle méthodologie de fusions.

Le présent rapport d'Audit du CESP porte sur le dispositif de l'année 2018 et plus précisément :

- le recrutement et la gestion du dispositif WEBALL, i.e. l'ensemble des individus qui participent à au moins un des trois panels
- l'analyse de la qualité des échantillons de chacun des trois panels
- l'examen de la méthodologie générale retenue par Médiamétrie pour fusionner les trois panels¹
- l'analyse de la stabilité des résultats d'audience

Au préalable, le Comité Scientifique souligne l'importance de la rédaction par Médiamétrie d'un corpus documentaire complet à destination du CESP du dispositif WEBALL, tant sur les contrôles métier appliqués en production que sur la méthodologie d'un point de vue technique. Compte tenu de la complexité du dispositif, cette documentation devra décrire précisément pour chaque étape de traitement les hypothèses, les données en entrée, les calculs réalisés et les résultats. Il sera également nécessaire de pouvoir disposer à la demande de données désagrégées et de codes informatiques pour certaines étapes de la modélisation.

Les étapes de pré-traitements conduites par Médiamétrie au niveau des panels Internet² n'ont pas été contrôlées et pourront faire l'objet de Focus complémentaires à la demande du marché. En particulier, le Comité Scientifique souligne l'importance de conduire deux Focus :

- Les redressements hybrides et leur impact sur les résultats d'audience (constitution des bases *site-centric*, règles de mise en cohérence des périmètres entre les données *site-centric* et les données mesurées par les panels, tests de sensibilité...).
- L'imputation de données https mise en œuvre par Médiamétrie depuis juin 2017 pour pallier les problèmes dans la collecte du surf sécurisé de panélistes qui utilisent Chrome comme navigateur (plus de 50% des internautes utilisent ce navigateur sur leur ordinateur³).

Un Focus sur l'estimation de la précision des résultats d'audience (à partir d'une approche par *bootstrap*) est programmé au premier semestre 2020.

Par ailleurs, 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, ce qui introduit un biais qui pourrait augmenter avec le temps si leur usage devenait plus répandu à l'avenir.

A partir de ses différents contrôles et analyses conduits sur l'année 2018, le CESP souligne que le « panel Trois Ecrans » à partir duquel les résultats d'audience sont désormais calculés est issu de fusions et correspond à un échantillon d'individus virtuels reconstitués. Il est construit principalement à partir du sous-échantillon des panélistes bi- et tri-mesurés,

¹ Le CESP a publié un focus spécifique en avril 2018 sur la nouvelle méthodologie des fusions.

² 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 « hybrides » est également hors du champ du Focus.

³ 52% en février 2018 d'après la source StatCounter (<http://gs.statcounter.com/browser-market-share/desktop/france>).

en moyenne 6 700 panélistes répondants actifs en 2018, qui sert de socle à l'élaboration des modèles de passage entre panels (les transformations procustéennes).

Le CESP considère satisfaisant :

- Le principe consistant à obtenir des résultats multi-*devices* à partir d'individus mesurés partiellement, compte tenu des difficultés à obtenir l'accord d'un panéliste pour être mesuré sur tous ses terminaux.
- L'augmentation de la part des recrutements de panélistes mobinautes et de panélistes bi- ou tri- mesurés.
- La qualité de la gestion des trois panels réalisés par l'équipe PRM : l'expertise des enquêteurs permet de répondre aux panélistes qui rencontrent des problèmes techniques, ce qui évite des abandons et favorise la participation à plusieurs panels.
- L'évolution à la baisse des taux de *churn* par panel depuis 2017. A titre d'exemple, sur le panel Mobile, le *churn* est passé de 13,8% en 2017 à 5,7% en 2018.

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 70 et 421 panélistes par strate en 2018).
- La forte variabilité dans la dispersion des poids de redressement du panel Ordinateur et du « panel Mobilité » due à l'introduction du nombre de visites *site-centric* d'un certain nombre d'entités comme critères de redressement.
- 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.

Le CESP recommande :

- De tester la possibilité de ne pas considérer dans chaque strate le Top 100 global des brands, mais le Top 100 de chaque strate afin de sélectionner les axes les plus pertinents pour chaque strate de fusion.
- D'augmenter la taille des sous-échantillons des panélistes bi- ou tri- mesurés pour améliorer les modèles de construction des fusions et de conduire des analyses complémentaires sur chaque strate afin d'apprécier l'hétérogénéité des profils et évaluer la robustesse de chaque modèle.
- De revoir la stratification définie *a priori* pour les fusions afin d'améliorer leurs performances.
- De ré-étudier l'intérêt de conserver les poids de redressement de chaque panel Internet. Les audiences par écran de référence pour le marché résultent du nouveau dispositif Internet Global et ne sont plus produites directement à partir de trois panels distincts. Lever cette contrainte permettrait d'envisager une autre approche méthodologique pour produire des audiences trois écrans.
- De fournir les effectifs bruts (nombre de panélistes réels) des 55 cibles standards et d'indiquer le seuil considéré comme minimum en termes d'effectif brut pour créer une cible dans l'interface.
- De justifier les tailles des différents panels et du nombre de bi-mesurés.

- De modifier le guide mis à disposition des utilisateurs dans l'interface de restitution des résultats, en précisant le calcul des seuils sur des individus virtuels reconstitués et non des panélistes bruts
- Concernant le panel Ordinateur :
 - De recueillir les habitudes de connexion à internet de la page de profil auprès de chaque panéliste, et non auprès d'un tiers, et ne plus avoir recours à une règle d'exclusion se basant sur l'incohérence observée entre les fréquences de connexion déclarées par un tiers et l'usage réellement mesuré⁴.
 - De faire chuter définitivement l'ensemble des panélistes identifiés par la règle d'exclusion sur la navigation atypique⁵ sur internet. En 2018, ces panélistes représentent en moyenne 1% du panel et sont le reliquat de recrutements *online* réalisés sur des bases d'adresses de moindre qualité.

Le Comité Scientifique souligne que les règles d'exclusion appliquées par Médiamétrie peuvent avoir un fort impact sur la taille et la représentativité des échantillons de panélistes contribuant aux calculs des audiences. En moyenne en 2018, 9% de l'échantillon du panel Ordinateur a été exclu.

- A la suite de la demande de Médiamétrie, le Comité Scientifique émettra un avis sur le sous-échantillon WORK dans son prochain rapport d'audit.

Le CESP recommande aux utilisateurs :

- 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 notifier que l'univers de mesure ne tient pas compte de certains systèmes d'exploitation.
- 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,
le 18 juin 2019*

⁴ 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.

⁵ La règles 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 appartenant aux deux sous-catégories « Jeux Concours / Loteries » et « Coupons de réduction ».