



C E S P

DISPOSITIF INTERNET GLOBAL

SYNTHESE DE L'AUDIT



2019

1

SYNTHESE DE L'AUDIT DU DISPOSITIF INTERNET GLOBAL 2019

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 méthodologie de fusions.

Le présent rapport d'Audit du CESP porte sur le dispositif de l'année 2019 et plus précisément :

- le recrutement et la gestion du dispositif, i.e. le suivi de 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 longitudinale des résultats d'audience

Ce rapport ne couvre pas l'ensemble des étapes de pré-traitements conduites par Médiamétrie au niveau de chaque panel Internet² et de la collecte des données qui pourront faire l'objet de Focus complémentaires. A la demande du Comité Scientifique, Médiamétrie travaille actuellement à la rédaction d'un corpus documentaire sur le dispositif Internet Global, 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, le Comité Scientifique souligne que 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 de données désagrégées et de codes informatiques pour certaines étapes de la modélisation. Sur la base de cette documentation exhaustive, le Comité Scientifique établira une liste hiérarchisée des Focus qui seraient à conduire. Il est d'ores et déjà prévu pour 2021 un Focus sur les redressements socio-démographiques et hybrides, sur les étapes de calage, 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...).

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 Audit. Le contrôle des redressements « hybrides » est également hors du champ.

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 versions d'OS 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 2019, 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, en moyenne 5 700 panélistes répondants actifs en 2019, 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.
- La qualité de la gestion des trois panels réalisée par l'équipe PRM (site Armatis) : 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.
- La progression des effectifs de panélistes bi- ou tri- mesurés pour la construction des modèles utilisés pour fusionner les panels Internet même s'ils restent insuffisants (entre 100 et 478 panélistes par strate en 2019).
- L'évolution à la baisse des taux de *churn* par panel depuis 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.
- 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.
- Le sous-échantillon des panélistes WORK en raison de sa non-représentativité de l'univers des individus ayant accès à internet sur leur lieu de travail. Le protocole de mesure mis en place (installation d'un *meter* sur un ordinateur professionnel non partagé), le mode de recueil des informations et l'absence de réponse aux questions sur la taille et le secteur d'activité de l'entreprise (respectivement 26% et 41%) conduisent le Comité Scientifique à s'interroger sur la qualité du recueil.
- La non communication des effectifs bruts (nombre de panélistes réels) des 55 cibles standards et l'absence de seuil minimum indiqué pour créer une cible dans l'interface.

- Le guide mis à disposition des utilisateurs dans l'interface de restitution des résultats qui ne précise pas que les résultats d'audience sont calculés sur des individus virtuels et non sur des panélistes bruts.

Le CESP recommande :

- De continuer à 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.
- 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 mesuré³.

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. Le CESP n'a pas disposé des résultats en 2019 qui lui permettent d'évaluer cet impact sur les effectifs du panel mais en moyenne en 2018, 9% de l'échantillon du panel Ordinateur avait été exclu.

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). A cet égard, le Comité Scientifique considère que l'utilisation du terme « panéliste » dans le guide est impropre.
- De notifier que l'univers de mesure ne tient pas compte de certains systèmes d'exploitation.

Le Comité Scientifique,
le 22 septembre 2020



³ 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.

Etat des lieux sur les points couverts par l'audit du CESP en 2019

Dans les tableaux ci-dessous le CESP a listé les principales étapes du dispositif Internet Global en précisant celles qui sont dans le champ⁴ de l'audit et celles qui ont été audités au 31 décembre 2019.

Dans les tableaux ci-dessous, nous avons indiqué par panel :

- En mauve : les points couverts par l'audit et contrôlés totalement ou partiellement
- En orange : les points couverts par l'audit 2019 mais non contrôlés
- En blanc : les points hors du champ de l'audit 2019

Panel Ordinateur			Objectif du contrôle CESP ou commentaire
Recrutement et gestion CATI	Champ	Contrôlé	
Recrutement et gestion CAWI (dont sous-échantillon WORK)	Champ	Non contrôlé en 2019. A contrôler en 2020	Emettre un avis sur l'ensemble de l'échantillon
Constitution des bases d'adresses	Champ	Partiellement contrôlé. A contrôler en 2020	Evaluer l'impact des sources de recrutement sur l'échantillon recruté
Relances	Champ	Contrôlé	
Recensement - CATI	Champ	Contrôlé	
Recensement - CAWI	Champ	Non contrôlé en 2019. A contrôler en 2020	En raison de la charge de son équipe, le CESP a fait le choix de ne pas couvrir ce point en 2019
Enquête Cible +	Hors champ		
Système de mesure : 1. Description 2. Suivi régulier des causes de limitation	Champ	Non contrôlé. Contrôler en 2020 le suivi des causes de limitations	Un audit du meter a été réalisé par le MRC. Le CESP demande à Médiamétrie d'apporter des éléments sur l'accréditation du MRC, notamment il s'agit de vérifier si l'audit est réalisé chaque année.
Système de mesure : 1. Traitements faits par Nielsen en amont sur les données remontée 2. Cas où les données remontées par le meter sont partielles ou manquantes, i.e. imputation chrome & Cas des panélistes « Outliers »	Hors champ	Non contrôlé en 2019. A contrôler et prévoir un Focus ultérieur	Vérifier si ces points sont dans le champ de l'audit du MRC. Apporter de la transparence sur le périmètre de la mesure et mettre en exergue les limitations
Incentives - Animation	Champ	Contrôlé	

⁴ L'ensemble des points couverts par l'audit sont discutés chaque année entre le CESP et Médiamétrie et des arbitrages sont décidés. Entrent dans le champ de l'audit les éléments listés dans l'Annexe Technique mise à jour chaque année.
22/09/2020

Panel Ordinateur			Objectif du contrôle CESP ou commentaire
Règles d'exclusion de panélistes	Champ	Partiellement contrôlé. A contrôler en 2020	
Redressement socio-démographique	Champ	Contrôlé	
Fusion Domicile et Lieu de travail	Hors champ	Non contrôlé en 2019. A contrôler et prévoir un Focus ultérieur si le sous-échantillon WORK est conservé	Examiner la méthodologie mise en œuvre (justification des hypothèses retenues, contrôle des éléments de validation) et évaluer l'impact sur les résultats
Elimination des panélistes « atypiques »	Hors champ	Non contrôlé en 2019. A contrôler et prévoir un Focus ultérieur	Evaluer la pertinence des règles définies par Médiamétrie (cf. Audit conduit en 2010) et leur impact sur la qualité de l'échantillon et sur les résultats
Redressement hybride* Calage post-redressement sur le nombre de pages vues <i>site-centric</i> *	Hors champ	Non contrôlé en 2019. A contrôler et à auditer en même temps que le Focus hybridation (en 2021)	Evaluer la pertinence de ce calage et son impact sur les résultats

* Ces points seront couverts dans le prochain Focus en 2021.

Panel Mobile & Panel Tablette			Objectif du contrôle CESP ou commentaire
Recrutement et gestion CATI	Champ	Contrôlé	
Recrutement et gestion CAWI	Champ	Non contrôlé en 2019. A contrôler en 2020	Emettre un avis sur l'ensemble de l'échantillon
Constitution des bases d'adresses	Champ	Partiellement contrôlé. A contrôler en 2020	Evaluer l'impact des sources de recrutement sur l'échantillon recruté
Relances	Champ	Contrôlé	
Recensement - CATI	Champ	Contrôlé	
Recensement - CAWI	Champ	Non contrôlé en 2019. A contrôler en 2020	En raison de la charge de son équipe, le CESP a fait le choix de ne pas couvrir ce point en 2019
Enquête Cible +	Hors champ		
Système de mesure : 3. Description 4. Suivi régulier des causes de limitation	Champ	Non contrôlé en 2019. A contrôler en 2020	Apporter de la transparence sur le périmètre de la mesure et mettre en exergue les limitations
Système de mesure - Traitements faits en amont sur les données remontées	Hors champ	Non contrôlé en 2019. A contrôler et prévoir un Focus ultérieur	
Incentives - Animation	Champ	Contrôlé	
Règles d'exclusion de panélistes	Champ	Partiellement contrôlé. A contrôler en 2020	
Redressement socio-démographique	Champ	Contrôlé	
Algorithme d'inférence développé par RealityMine pour identifier les applications étant en focus dans le cas d'un smartphone iOS 8 ou plus (Panel Mobile uniquement)	Hors champ	Non contrôlé en 2019. A contrôler et prévoir un Focus ultérieur	Evaluer la qualité de cet algorithme et l'impact sur l'audience
Modélisation mise en œuvre par Médiamétrie pour affecter les sessions de navigation à chacun des membres du foyer (Panel Tablette uniquement)	Hors champ	Non contrôlé en 2019 mais sera intégré dans l'audit 2020 compte du changement de système de mesure du panel Tablette	Analyser la modélisation mise en œuvre et disposer d'éléments de validation sur l'affectation correcte du surf aux individus d'un même foyer
Redressement hybride* Calage post-redressement sur le nombre de pages vues <i>site-centric</i> *	Hors champ	Non contrôlé en 2019. A contrôler et à auditer en même temps que le Focus hybridation (en 2021)	Evaluer la pertinence de ce calage et son impact sur les résultats

* Ces points seront couverts dans le prochain Focus en 2021.