

Metodbeskrivning

MMS Räckviddsmodell för Online + Total TV





Innehåll

Inledning	3
Steg 1: Profilering	3
1.1 Censusedata.....	3
1.2 Paneldata	4
1.3 Skapa prediktionsmodell	4
1.4 Applicera prediktionsmodell.....	6
1.5 Modellering av målgrupp 3-14	9
Steg 2: Beräkning av räckvidd online	10
2.1 Unika cookies per målgrupp	11
2.2 Förhållandet mellan cookies och användare	11
2.3 Populationstal.....	11
2.4 Plattformsspecifik räckvidd	11
2.5 Populationstal Cross-Device	12
2.6 Korrelation mellan plattformar	12
2.7 Beräkning av räckvidd Online	14
Steg 3: Kombinera online med TV	15
3.1 Plattformsspecifik räckvidd	15
3.2 Korrelation mellan TV och Online-plattformar.....	16
3.2 Beräkning av räckvidd Online+TV	17

Inledning

Maj 2017 lanserade MMS räckviddsmätning i olika demografiska målgrupper för reklamfilmer som har visats online på uppmätta videotjänster. Detta redovisas i MMS analysverktyg Reklamanalys. I september 2017 kompletterades mätningen med möjligheten att söka fram den plattformsovergripande räckvidden för reklamkampanjer som har visats både online och på TV.

Detta dokument syftar till att beskriva metodiken bakom dessa mätningar. För användarinformation om Reklamanalys och de olika mått som redovisas i verktyget hänvisas till [Användarinfo Reklamanalys](#).

Från och med maj 2018 finns också möjlighet att i analysverktyget Programanalys fördela måtten starter, webbpublik och summa tid i samma demografiska målgrupper. Detta bygger på den profileringsmodell som beskrivs i avsnitt 1.1-1.5 i detta dokument. En räckviddsmodell för att kunna presentera räckviddsmått för program och klipp är under utveckling.

Steg 1: Profilering

1.1 Censusedata

MMS mäter all reklamfilm och program/klipp som visats på de mätta videotjänsterna, där varje film/titel har en unik mättagg som fångas upp av MMS i alla de olika spelare och appar som mäts. Mätdata skickas till MMS som kvalitetskontrollerar, sammanställer och publicerar den officiella trafiken för både reklam och program/klipp. Från denna mätning ges för reklamen information om den totala volymen av impressions och completion rate för samtliga reklamfilmer. För program och klipp möjliggör censusedata redovisning av populationens totala antal starter och total spenderad tid på titelnivå.

I censusedata fångas upp information om cookie eller Ad-ID (vid konsumtion via app på mobila enheter) bakom videostarten. Detta innebär att för varje unik cookie finns en totalbild av dess konsumtion av såväl reklamfilm som program & klipp på de olika uppmätta siter. Dock saknas information om individerna bakom konsumtion. För att möjliggöra redovisning av mått inom olika sociodemografiska målgrupper krävs därmed ytterligare datakällor samt statistisk modellering.





1.2 Paneldata

En nödvändig källa för att möjliggöra räckviddsmätning är paneldata, där det finns sociodemografisk bakgrundsinformation om en uppsättning individer vars trafik på de uppmätta siterna kan identifieras via cookies.

MMS använder sig av paneldata bestående av totalt ca 25 000 rapporterande individer, varav ca 5000 utgör den så kallade kärnpanelen. Kärnpanelen ska vara representativ för Sveriges Internetpopulation och omfattas av en daglig validering gentemot en uppsättning regler som syftar till att få en så representativ, kvalitativ och heltäckande kärna som möjligt.

Panelist	Kön	Ålder
1	Man	19
2	Kvinna	35
3	Man	55
4	Man	27
...
...
N	KönX	ÅlderY

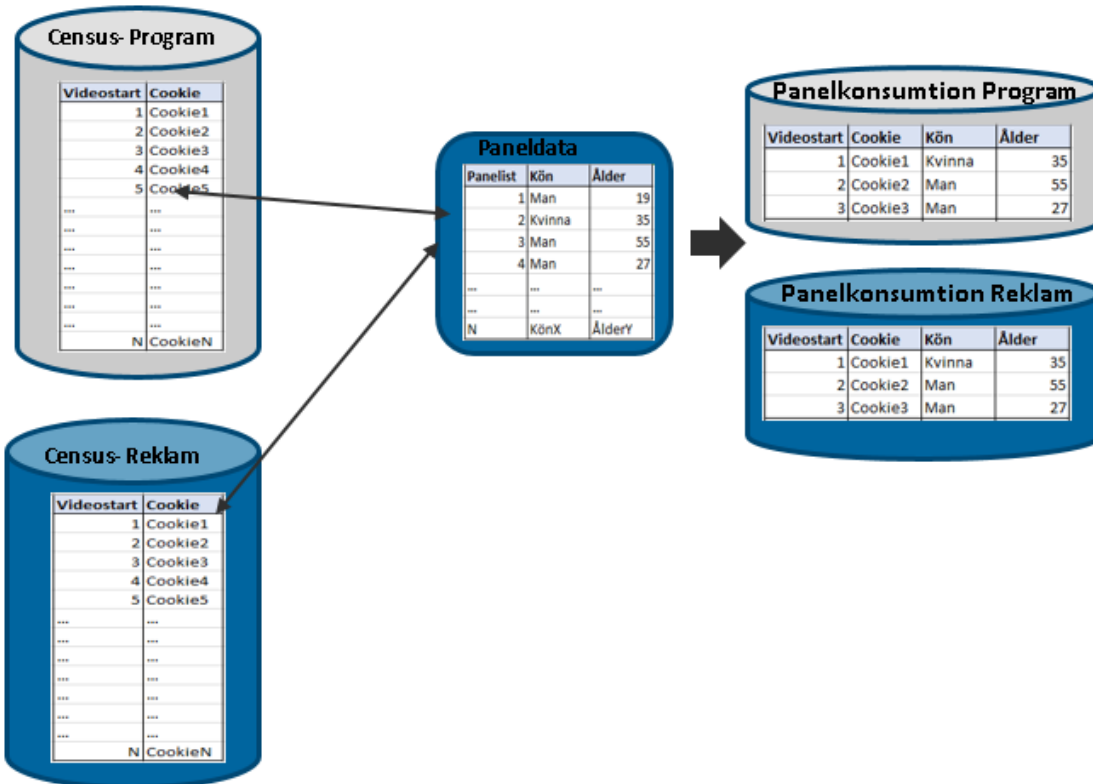
Panelisterna ombeds registrera samtliga enheter (dator, smartphone, surfplatta) som de använder internet på, och i varje browser på respektive enhet ska de placera en cookie som möjliggör att deras konsumtion kan identifieras i censusmätningen. För identifiering av konsumtion via app (på mobila enheter) behöver panelisten installera en särskild panel-app som möjliggör ihopkoppling mellan panelistens registrerade enhet och trafikdata i census.

1.3 Skapa prediktionsmodell

Att rakt av använda en onlinepanel och vikta upp dess uppmätta konsumtion för att representera befolkningens är problematiskt på grund av den stora fragmentisering online. Många objekt (titlar/reklamkampanjer) med låg räckvidd i befolkningen kräver en väldigt stor panelstorlek för att kunna mäta mer än endast de allra största titlarna/reklamfilmerna.

En stor panelstorlek medför dock andra svårigheter som att rekrytera tillförlitliga panelister och upprätthålla representativitet avseende både sociodemografi som fördelning av olika kombinationer av använda enheter.

Därför är det nödvändigt med kompletterande metodik för att kunna utnyttja den styrka som en kvalitativ kärnpanel ger men samtidigt möjliggöra rapportering ner på objekt med få antal starter. MMS metod startar med en profilering på cookienivå, där identifierad konsumtion från hela panelen (25 000 individer) ligger till grund för att skapa en prediktionsmodell.



Genom panelisternas konsumtion har vi för en delmängd av trafiken i censusdata information om exempelvis kön och ålder för de individer som genererat konsumtionen. Denna datamängd används för att analysera konsumtionsmönster med syfte att hitta samband mellan programtitlar/reklamfilmer och de olika målgrupperna.

Detta görs dagligen och löpande används den sammanlagda panelkonsumtionen från de senaste 28 dagarna för att bilda ett omfattande underlag till modelleringen. För kombinationerna av titel/reklamfilm och målgrupp skattas sedan en koefficient, där ett högt värde på koefficienten innebär ett starkt observerat samband mellan den aktuella titeln och målgruppen.

Resultatet är en prediktionsmodell som beskriver sannolikheten för att tillhöra en specifik målgrupp:

$$Sannolikhet(Målgrupp) = Koeff_1 * Titel_1 + Koeff_2 * Titel_2 + \dots + Koeff_n * Titel_n$$



1.4 Applicera prediktionsmodell

Prediktionsmodellen används sedan till att skatta demografiska profiler på alla okända id:n i censusdata, baserat på den konsumtion som observerats för respektive cookie. Precis som i föregående steg används den sammanlagda trafiken från de 28 senaste dagarna för varje cookie.

Om en cookie har konsumerat titlar som har observerats ha ett starkt samband med en viss målgrupp, så medför detta att sannolikheten för att cookien tillhör den aktuella målgruppen ökar.

Det går aldrig att med 100% säkerhet att säga att en cookie tillhör en viss profil, och om cookien enbart har konsumerat innehåll som i panelen visat sig attrahera en väldigt bred publik så kan det vara svårt att t.ex särskilja mellan olika åldersgrupper.

Därför är resultatet av profileringen inte att varje cookie får en utpekad specifik profil, utan istället ges en sannolikhet för var och en av de olika målgrupperna:

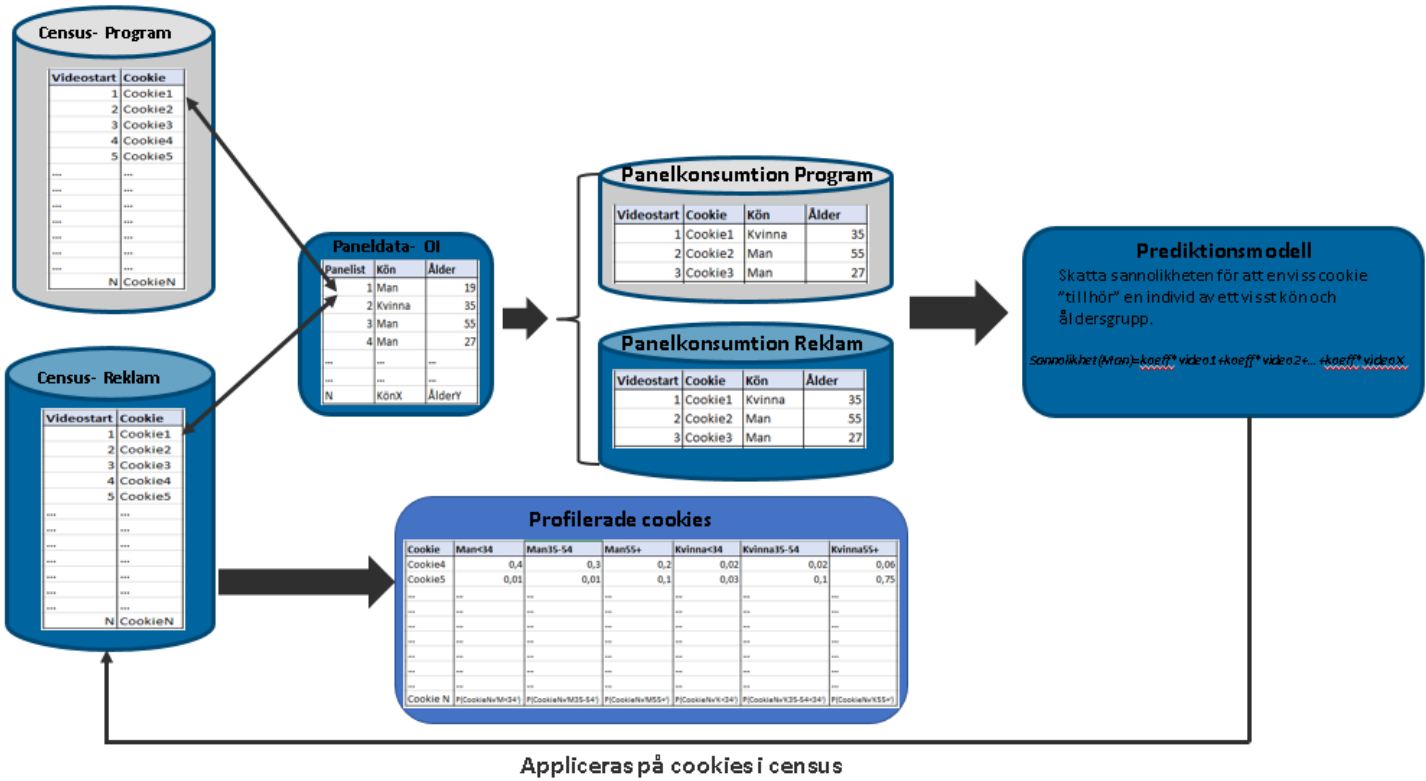
Cookie	Man<34	Man35-54	Man55+	Kvinna<34	Kvinna35-54	Kvinna55+
Cookie4	0,4	0,3	0,2	0,02	0,02	0,06
Cookie5	0,01	0,01	0,1	0,03	0,1	0,75
...
...
...
...
...
...
...
Cookie N	$P(\text{CookieN}=\text{M}<34)$	$P(\text{CookieN}=\text{M}35-54)$	$P(\text{CookieN}=\text{M}55+)$	$P(\text{CookieN}=\text{K}<34)$	$P(\text{CookieN}=\text{K}35-54<34)$	$P(\text{CookieN}=\text{K}55+)$

Från och med maj 2018 profileras 18 olika målgrupper i profileringsmodellen, dessa är:

Målgrupp	
Män 3-14	Kvinnor 3-14
Män 15-19	Kvinnor 15-19
Män 20-24	Kvinnor 20-24
Män 25-29	Kvinnor 25-29
Män 30-44	Kvinnor 30-44
Män 45-49	Kvinnor 45-49
Män 50-59	Kvinnor 50-59
Män 60-64	Kvinnor 60-64
Män 65-99	Kvinnor 65-99



När profilerat censusdata sedan används för att möjliggöra fördelning av antalet impressions per målgrupp, så ger tillvägagångssättet med sannolikheter ett mer nyanserat och rättvisande resultat jämfört med om varje cookie skulle kopplas samman med enbart den målgrupp där sannolikheten är högst.

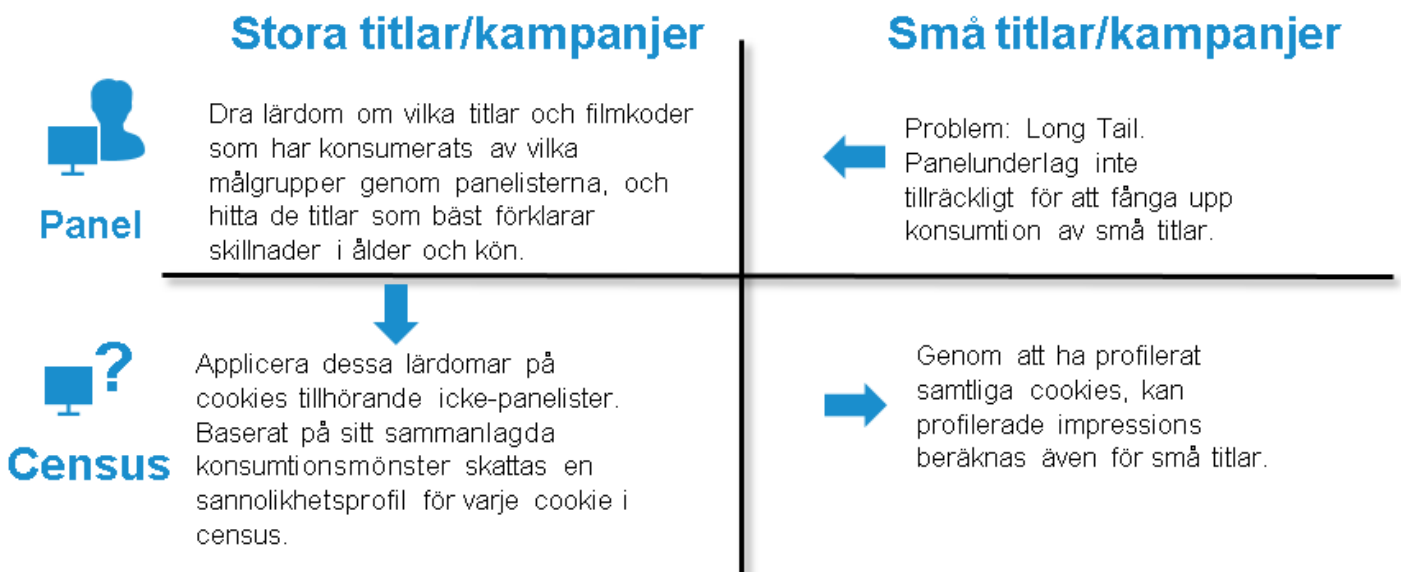




Principen för profileringen kan också beskrivas med hjälp av nedanstående bild där vi grovt delat in mätobjekten i stora respektive små titlar/kampanjer.

Som tidigare konstaterat är problemet med en rent panelbaserad mätning att panelunderlaget inte är tillräckligt för att på ett tillförlitligt sätt fånga upp konsumtionen på små objekt. Däremot kan vi för de tillräckligt stora titlarna/filmkoderna dra lärdomar av vilka målgrupper som konsumerat dessa i panelen.

Det är denna kunskap som används för att skapa prediktionsmodellen som sedan appliceras på cookies tillhörande icke-panelister, och tack vare att vi profilerar samtliga cookies i census kan vi överföra kunskapen även på små titlar/kampanjer genom vetskapen om vilka cookieprofiler som konsumerat dem.





1.5 Modellering av målgrupp 3-14

Åldersgruppen 3-14, den så kallade barnmålgruppen, är speciell då onlinepanelen inte inkluderar några panelister med denna ålder. Det är därmed med hjälp av panelen inte möjligt att lära sig något om hur denna grupps konsumtionsmönster ser ut.

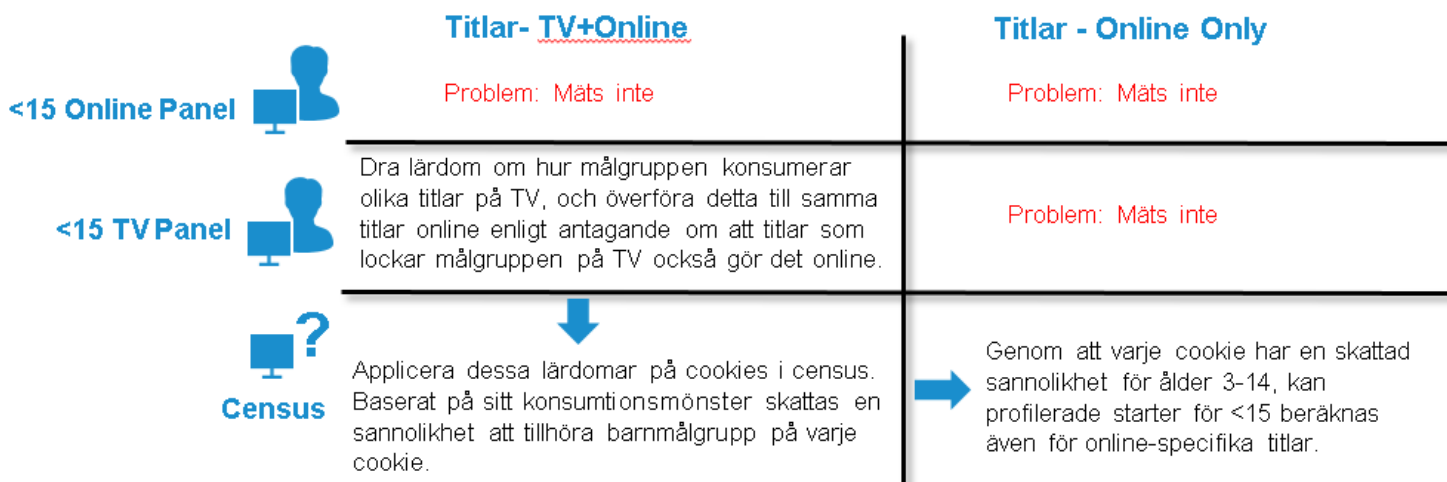
För att kunna skatta sannolikheten att en cookie är 3-14 år och därmed möjliggöra utbrytning av denna målgrupp i resultatet, har MMS behövt skapa en specialmodell just för denna åldersgrupp. En första version av denna "barnmodell" lanserades maj 2018 i samband med att antalet profilerade målgrupper utökades från 6 till 18 stycken. En viktig input i modellen är målgruppens konsumtion på TV vilket mäts via MMS TV-panel.

Nedan bild visualiserar problematiken på ett liknande sätt som tidigare, men där titlarna istället delats upp beroende på om de är online-specifika eller om de visas både via TV och online video.

Som redan konstaterats så mäts åldersgruppen inte alls via online-panelen, och det online-specifika materialet kommer heller inte att fångas upp av TV-panelen. Däremot ger TV-panelen information om hur målgruppen konsumerar olika titlar på TV. Genom antagande om att titlar som lockar målgruppen på TV också gör det online, kan vi överföra dessa lärdomar som input till prediktionsmodellen.

Detta möjliggör att vi för varje cookie kan skatta sannolikheten även att tillhöra "barnmålgruppen". Principen för detta kan förenklat beskrivas som att om en cookie har observerats konsumera titlar med hög andel TV-räckvidd i gruppen 3-14 år, så kommer detta att bidra till en ökad sannolikhet för att cookien tillhör denna målgrupp.

När alla cookies har profilerats kan vi beräkna profilerade starter för målgruppen även på online-specifika titlar (och även reklamfilmer), ekvivalent med hanteringen av små titlar/kampanjer.



Som tidigare nämnt är denna modell en första version, och MMS kommer att fortsätta att arbeta med att vidareutveckla barnmodellen för att göra den så robust och rättvisande som möjligt. Det finns en medvetenhet om att modellen bygger på ett relativt grovt antagande, och att räckvidder för samma titel kan skilja sig mellan TV och via online plattformarna. Aspekter som tillsammanstittande är också en faktor som försvagar antagandet.

I den första versionen skattas sannolikheten totalt för åldersgruppen 3-14, och fördelas sedan 50/50 på flickor och pojkar. Ett framtida utvecklingsbehov är därmed att bättre differentiera mellan könen i barnmålgruppen.

Steg 2: Beräkning av räckvidd online

Profileringen av censusdata möjliggör fördelning av totalt antal impressions/starter på olika målgrupper, men är inte tillräcklig för att redovisa räckviddsmått i form av unika användare.

Det finns flera kända problem med cookiedata:

- En cookie kan tillhöra flera olika individer (vid delade enheter).
- En individ kan ha flera olika cookies (vid användning av olika enheter, eller flera olika browsers alt. browser+app på en enhet.)
- Cookies är inte beständiga över tid utan kan raderas och förnyas.

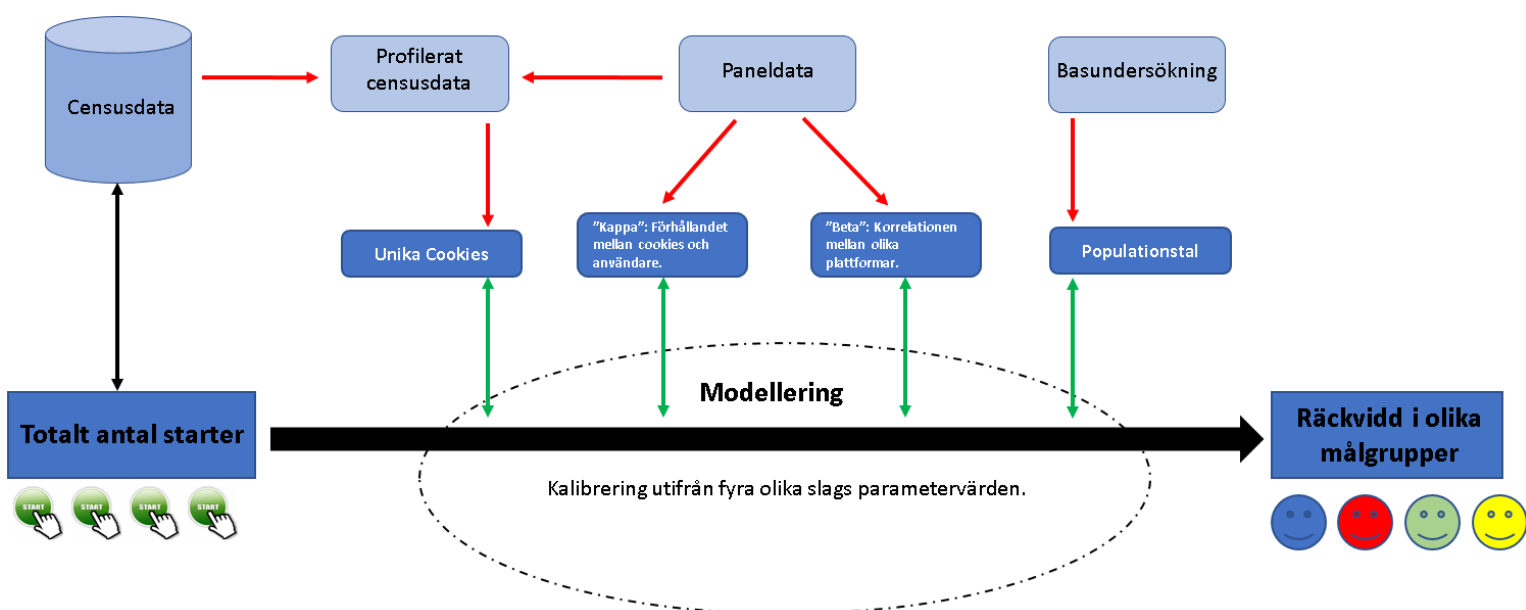
För att möjliggöra beräkning av räckvidd behöver ovan nämnda aspekter hanteras, vilket kräver ytterligare modellering.

MMS har valt en metod som tar utgångspunkt i profilerat censusdata, och använder sedan kärnpanelen och populationstal för internetanvändning på olika typer av enheter för att skatta olika nödvändiga parametrar för kalibrering av räckviddsmodellen.

Utöver att härleda parametrar till modellen så har kärnpanelen en viktig roll för validering av modellens resultat. Tanken är att räckvidden, för kampanjer som är tillräckligt stora för att kunna mätas genom en ren panelbaserad mätning, ska vara densamma både från räckviddsmodellen samt direkt från panelen, samtidigt som modellen tillåter förlängning av detta även till de betydligt mindre kampanjerna.

Modellen använder fyra olika slags parametervärden som kommer att gås igenom i de följande avsnitten:

- 1) Unika cookies per målgrupp och device
- 2) Förhållandet mellan cookies och användare ("kappa")
- 3) Korrelationen mellan olika plattformar ("beta")
- 4) Storleken på populationen





2.1 Unika cookies per målgrupp

Utgångspunkten i räckviddsmodellen är antalet unika cookies per devicetyp och målgruppsprofil. Källan för detta är den profilering som beskrivs under avsnitt 1.3 och 1.4.

Om en cookie som observerats på desktop får sannolikheten 0,6 att tillhöra exempelvis "blå" målgrupp så innebär det att denna bidrar med 0,6 cookies när vi beräknar alla målgruppens cookies.

2.2 Förhållandet mellan cookies och användare

Som tidigare konstaterat är unika cookies inte detsamma som unika användare. Vi behöver därför kunskap kring hur antalet cookies relaterar till antalet användare.

Till detta använder vi MMS kärnpanel där vi har information om individer och deras målgruppsinformation, alla deras olika enheter samt deras olika cookies på respektive enhet och browser över tid. Utifrån det observerade mönstret i panelen så estimeras parametern "kappa" för att kalibrera förhållandet mellan cookies och användare.

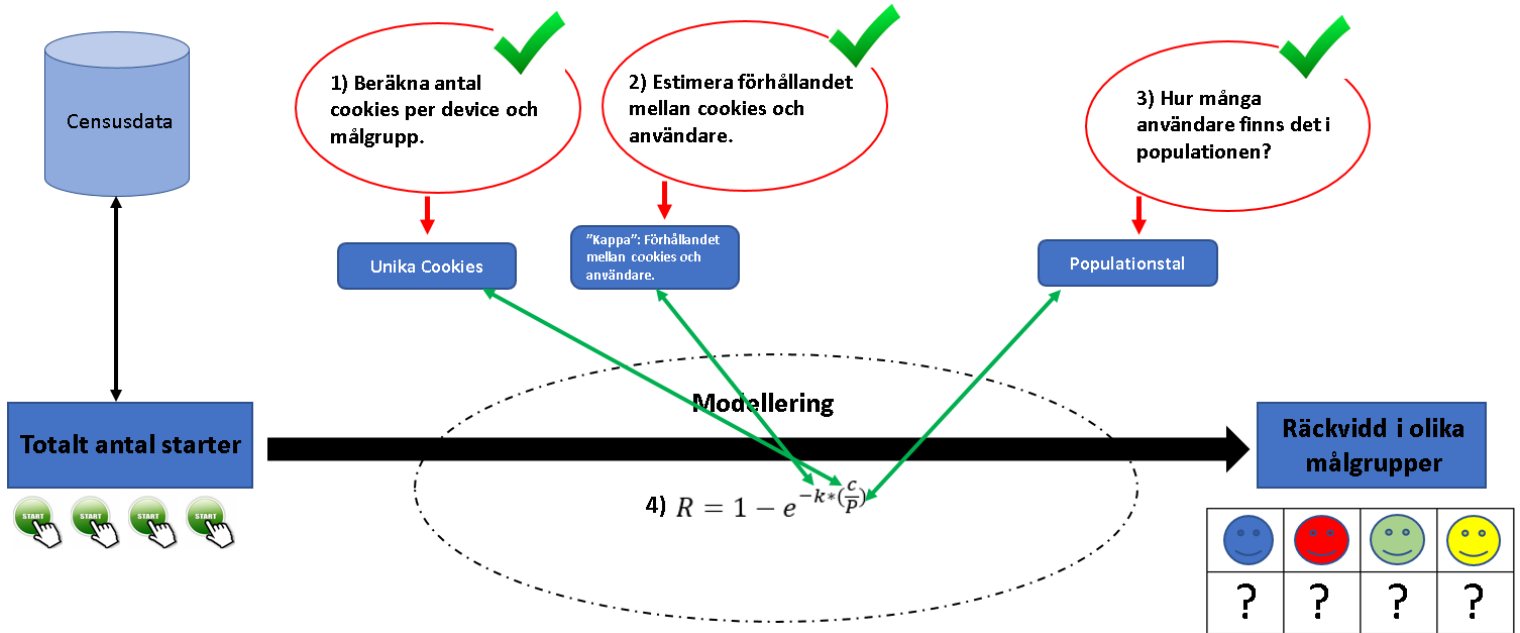
Respektive plattform har sin egen kappa eftersom cookiemönstret ser olika ut på olika slags plattformar. Kappa utgör ett värde mellan 0 och 1 och blir lägre desto fler cookies per individer, t ex om större förekomst av cookieradering eller antal olika webbläsare per individ.

2.3 Populationstal

Modellen behöver också ta hänsyn till antalet potentiella användare i populationen, dvs antalet internetanvändare per plattform i respektive målgrupp. Källan för dessa populationstal är MMS Basundersökning. De populationstal som används i modellen kommer att uppdateras två gånger per år, i samband med TV-mätningens universumtalsskifte.

2.4 Plattformsspecifik räckvidd

Utifrån parametrarna från steg 1-3 kan nu det första beräkningssteget i räckviddsmodellen genomföras:



Den specifika räckvidden ("R") beräknas inom varje kombination av plattform och målgrupp:

$$R = 1 - e^{-k * (\frac{c}{P})}$$

Där c=Antal unika cookies, k=Kappa och P=Populationstal.

2.5 Populationstal Cross-Device

För att kunna beräkna den slutgiltiga räckvidden behövs också tas hänsyn till beroende mellan plattformarna. En nödvändig parameter i detta är mer detaljerade populationstal över hur de olika målgrupperna använder olika kombinationer av plattformar/device-typer.

Det som behövs är storleken på målgruppen som t.ex använder enbart dator, som använder både dator och smartphone, som använder på dator+smartphone+tablet etc.

Precis som i punkt 2.3 är källan MMS Basundersökning.

2.6 Korrelation mellan plattformar

Utöver kalibrering för antal potentiella användare i populationen behöver också hänsyn tas till eventuellt beroende i konsumtionen mellan plattformar. T.ex:

- Om man har sett en reklamfilm på en plattform, påverkar det sannolikheten för att man ser samma reklamfilm på en annan plattform?
- Hur benägen är man att konsumera något på flera olika plattformar?

Precis som i estimeringen av Kappa i punkt 2.2 så används kärnpanelens konsumtionsmönster för att lära om beroendet mellan plattformar. I denna parameterestimering är det av väldigt stor vikt att panelen mäts

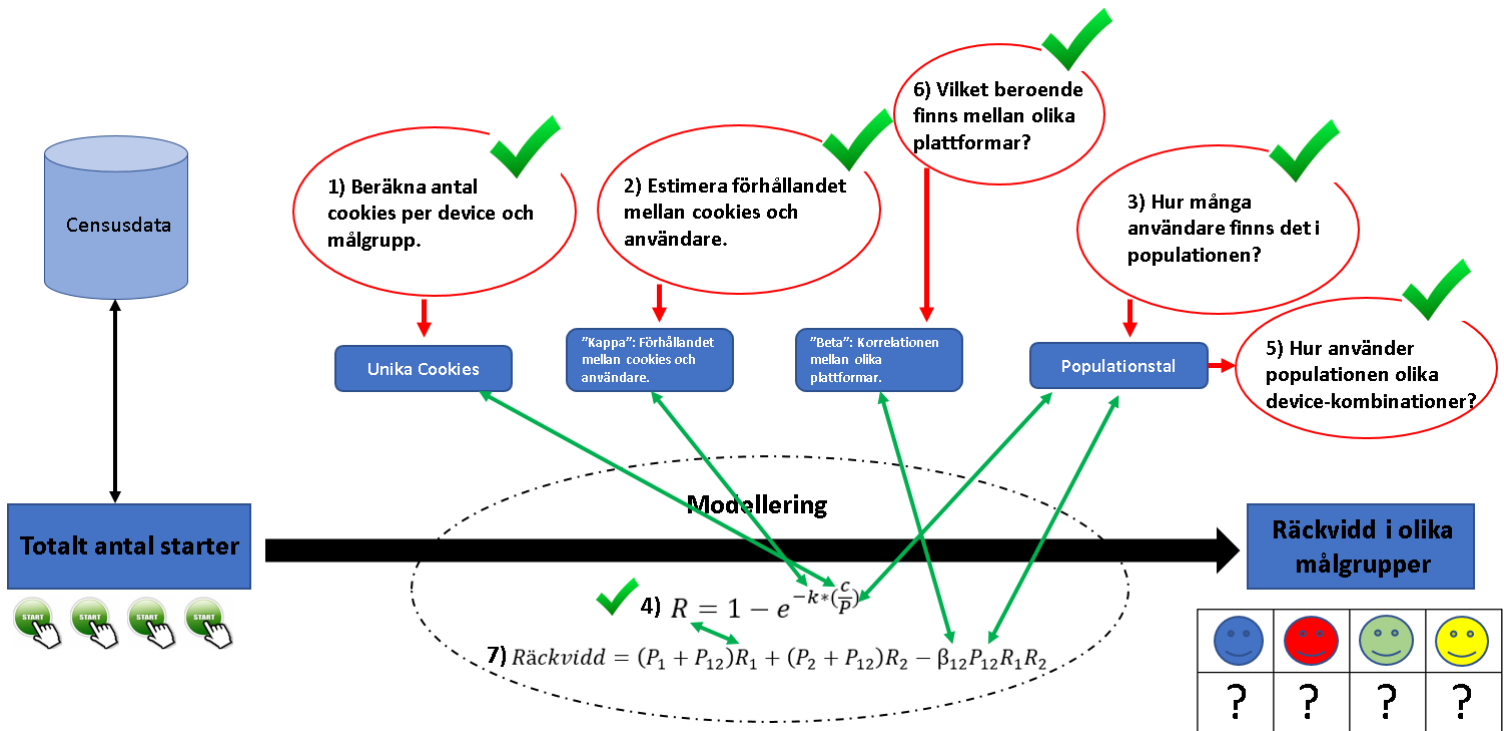


cross-device, dvs över alla sina använda enheter, för att möjliggöra observation av individers konsumtion på olika plattformar.

Utifrån panelens observerade konsumtion på olika plattformar estimeras parametern β . $\beta = 1$ om räckvidden på två plattformar har observerats vara oberoende av varandra. $\beta > 1$ om positiv korrelation och < 1 om negativ korrelation mellan plattformarnas räckvidd.

2.7 Beräkning av räckvidd Online

Efter stegen 2.1-2.6 finns alla nödvändiga komponenter för att beräkna den slutliga plattformsovergripande räckvidden i olika målgrupper.



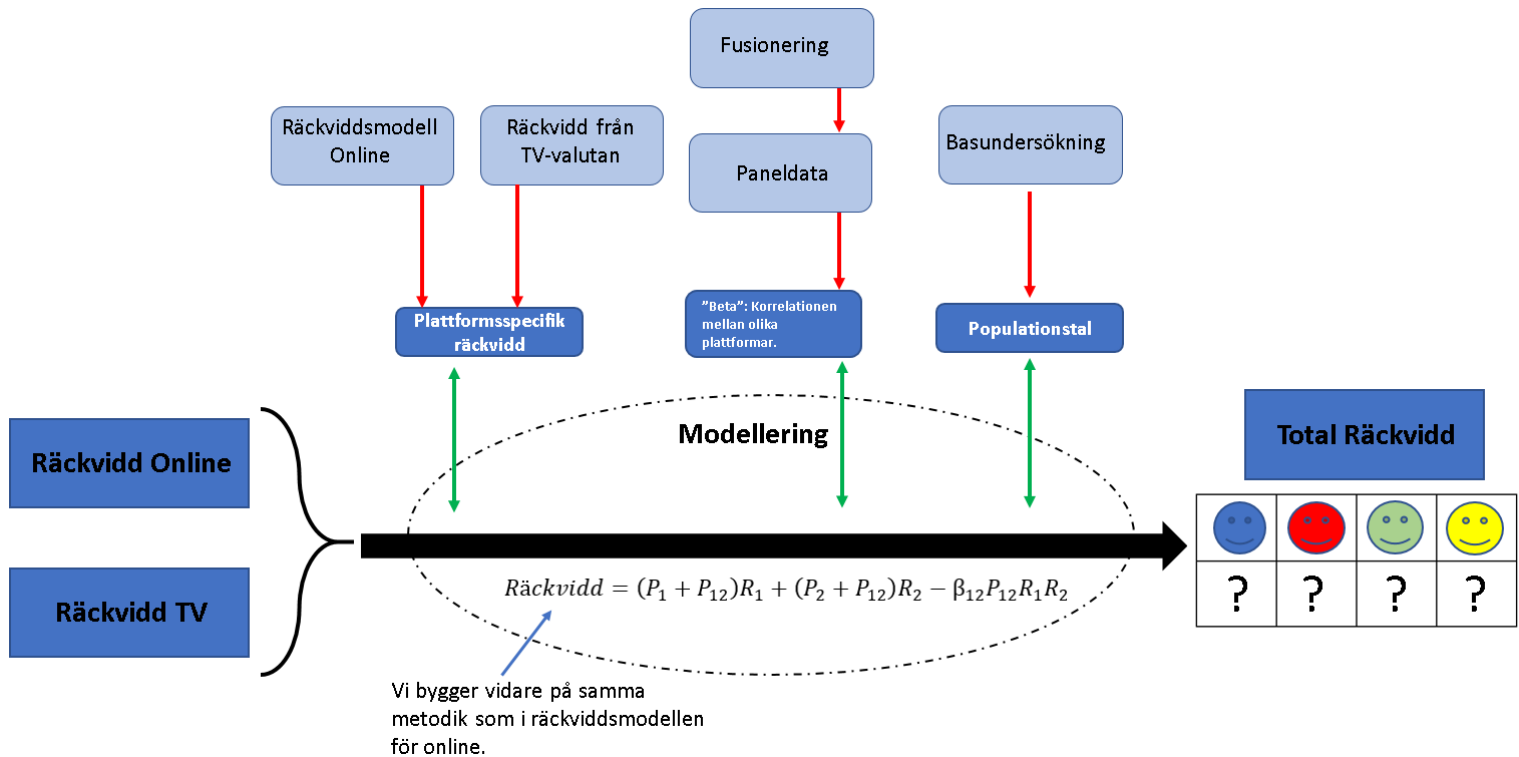
För ett exempel med två olika plattformstyper (1 och 2) sker beräkningen för en målgrupp med formeln:

$$Unika\ användare_{12} = (P_1 + P_{12}) * R_1 + (P_2 + P_{12}) * R_2 - \beta_{12} * P_{12} * R_1 * R_2$$

Där P_1 är antalet i den specifika målgruppen som använder enbart device 1, P_2 är antalet som använder enbart device 2 och P_{12} är de som använder båda devicetyperna. R är den specifika räckvidden för respektive plattform som beräknades i steg 2.4

Steg 3: Kombinera online med TV

Den totala räckvidden för kampanjer som visats både på TV och någon av de mätta online-siterna bygger vidare på ovan beskriven metodik, där TV introduceras som ytterligare en plattformstyp.



3.1 Plattformsspecifik räckvidd

Första steget är precis som för onlineräckvidden att beräkna den plattformsspecifika räckvidden per varje kombination av målgrupp och devicekombination, dvs parametern "R".

För online-plattformarna sker detta på exakt samma sätt som i avsnitt 2.4, där unika cookies transformeras till unika användare genom kalibrering mot populationstal och parametern kappa. Detta beräkningssteg är varken applicerbart eller nödvändigt för TV-räckvidden. Istället används räckvidden för TV per målgrupp beräknad som i TV-mätningen.

När räckvidden på TV beräknas över en period så utgår alltid beräkningen ifrån en så kallad *paneldag*, som per default utgör mittendagen i perioden (den övre av de två mittendagarna vid jämnt antal dagar i perioden.). Det innebär att man utgår från panelens sammansättning och viktning den aktuella paneldagen, och beräknar sedan räckvidden per dag utifrån den specifika panelsammansättningens konsumtion över hela perioden.

En konsekvens av detta är att en sökning på period dag 1-6 respektive dag 1-7 kan ge olika resultatet i och med att olika paneldagar kommer att ligga till grund för beräkningen.

För mer information om TV-valutans mått och beräkningsregler hänvisas till [MMS Golden rules](#).

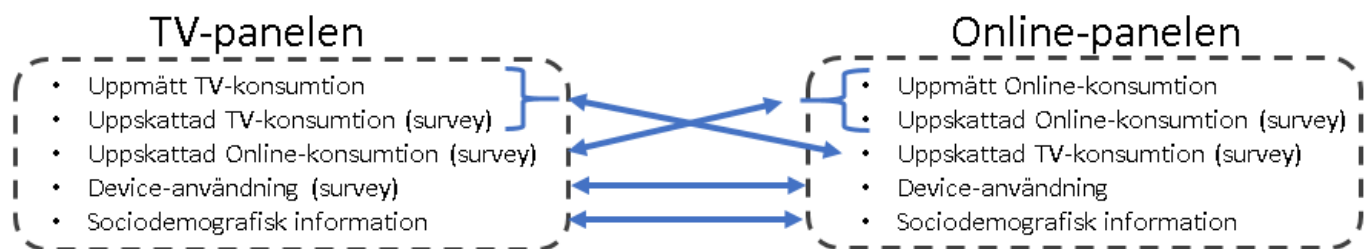
3.2 Korrelation mellan TV och Online-plattformar

Återigen behöver vi identifiera vilket beroende som finns mellan konsumtion på TV respektive konsumtion på online-plattformarna för att kunna beräkna den totala räckvidden.

Precis som i steg 2.6 då vi estimerade parametern β i räckviddsmodellen för Online, så är detta något vi kan lära oss genom att studera konsumtionsmönster i en panel som mäts över samtliga plattformar. Kruxet är dock att TV- och Online-konsumtionen inte observeras i samma panel, utan mäts i två olika.

För att kunna observera individers konsumtion över samtliga plattformar och därifrån kunna estimeras β till den totala räckviddsmodellen, så används fusioneringsmetodik för att sammanföra TV-panelen och Online-panelen till en total panel.

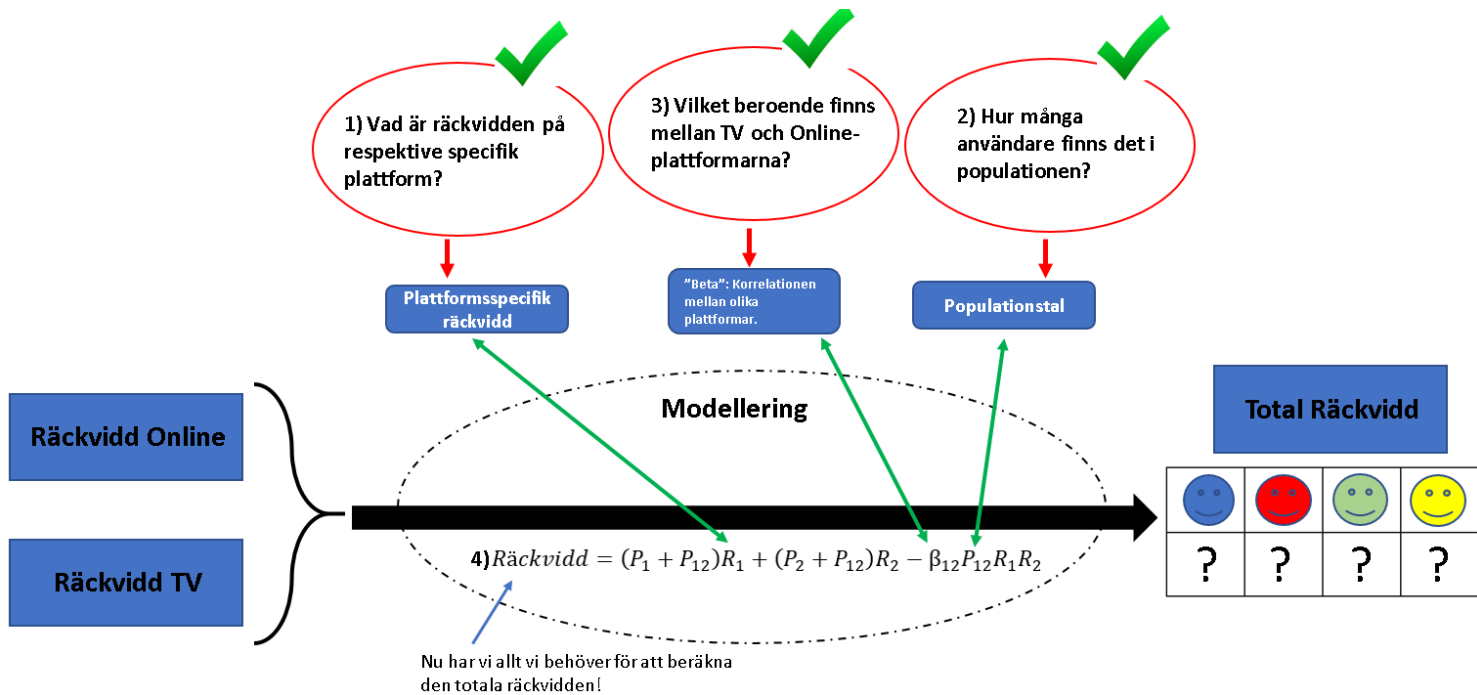
Fusionering handlar om att via statistisk metodik matcha ihop panelister från de olika panelmängderna. De panelister som paras ihop liknar varandra givet en sammansättning av olika variabler. De variabeltyper som används i fusioneringen är sociodemografisk bakgrundsinformation, självuppskattad och uppmätt TV-konsumtion, självuppskattad och uppmätt Online-konsumtion samt uppgiven device-användning.



Fusioneringen sker med utgångspunkt i TV-panelen och dess viktning. En TV-panelist som matchas med en Online-panelist kommer i de fusionerat paneldata att behålla sin TV-konsumtion men samtidigt också tilldelas onlinekonsumtionen från Online-panelisten. En TV-panelist kan fusioneras med exempelvis två olika online-panelister, och kommer då att utgöra två panelister med halverad ursprungsvikt i den fusionerade panelen.

När fusioneringen är genomförd har vi en total panelmängd som innehåller såväl TV- som Online-konsumtion, och detta kan användas för att observera i vilken utsträckning individer konsumerar på både TV och andra plattformar. Precis som i 2.6 kan vi nu estimeras parametern β med TV-plattformen inkluderat.

3.2 Beräkning av räckvidd Online+TV



Det är återigen formeln från avsnitt 2.7 som är aktuell, där TV alltså inkluderas som ytterligare en plattform.

Nödvändiga parametrar i beräkningen av total räckvidd är därmed:

- R=plattformsspecifik räckvidd, där R för TV beräknas utifrån TV-panelen och TV-valutans beräkningsregler för räckvidd. R för övriga plattformar beräknas enligt steg 2.4
- P=populationstal för respektive målgrupps- och devicekombination. T.ex: Hur många har TV men använder ingen online-plattform? Hur många har både TV och använder internet på smartphone? Precis som i tidigare steg används populationstal från MMS basundersökning för detta.
- β =Observerat beroende mellan plattformarna. Denna parameter härleds ur en fusionerad panelmängd där individers konsumtion över både TV och Online kan observeras.

Populationstalen behöver justeras två gånger per år i samband med ordinarie universumtalsbyte för TV-valutan. Även de estimerade parametrarna kappas och β behöver kontrolleras och eventuellt justeras/omkalibreras med jämna mellanrum, och även detta är planerat att sammanfalla med universumtalsbyte. Vid omestimering av β görs även fusioneringen om på en nyare datamängd.