

Modellbeskrivning

MMS Räckviddsmodell för Onlinevideo





Innehåll

Inledning	3
Historik	4
1 Datakällor	5
1.1 Censusedata	5
1.2 Paneldata.....	6
2 Modellering	7
2.1 Imputering.....	7
2.2 Profilerings.....	8
2.3 Kloning.....	10
2.4 Cookie deletion-modell.....	12
3 Den virtuella panelen	13
3.1 Panelsammansättning	13
3.2 Viktning och universumtal.....	13
3.3 Kalibrering mot census.....	15
3.4 Beräkning av publikmått.....	16
4 Appendix	17
4.1 Big Screen	17
4.2 Barnmålgrupp	18
4.3 Kloning av barnmålgrupp	19



Inledning

Från och med maj 2021 (för Program och klipp) och 2022-09-26 (Reklam) redovisar MMS räckviddsmått, och andra publikmått, som beräknas utifrån MMS virtuella panel. Metodiken bakom den virtuella panelen går också under namnet AAM ("audience ascription model") och bygger på en kombination av i huvudsak två datakällor:

- Censurmätningen där trafikvolymerna mäts på detaljerad nivå ända ner till enskilda starter, men som på egen hand inte ger någon information om den publik som genererat trafiken.
- Videopanelen där 3000 hushåll mäts på såväl TV-skärmen som sina digitala enheter och ger information om hur olika målgrupper konsumerar, men som på egen hand inte är tillräcklig för att kunna täcka den långa svansen av fragmentiserat tittande.

Dessa kombineras till en gemensam källa i form av ett paneldataset bestående av ca 46 000 virtuella panelister per dag, och deras tittande från censurmätningen (över såväl smartphone, surfplatta, PC och den stora skärmen). Detta dataset har ett individperspektiv där varje panelist är viktad för att representera ett visst antal individer i befolkningen. Därmed kan räckvidd och andra publikmått beräknas på liknande villkor som i en rent panelbaserad modell (exempelvis MMS TV-mätning), och möjliggör kommande integrering med TV-mätningen för en total plattformsovergripande mätning i en gemensam panel.

Syftet med den virtuella panelen är att kombinera styrkorna från respektive datakälla. Den utökade panelstorleken ger en betydligt större täckningsgrad och därmed möjligheter att analysera publikmått på mindre titlar än vad enbart onlinepanelen skulle tillåta. Resultatet från den virtuella panelen kalibreras också för att så långt det är möjligt redovisa samma totalvolymerna av starter och tid som uppmätt i censurmätningen.

Processen för att skapa den virtuella panelen inkluderar ett antal olika modelleringssteg så som:

- Imputering av icke-mätta enheter i panelen
- Kloning av panelister, inklusive modell för cookie deletion
- Kalibrering mot censurnivåer
- Viktning av panelister

Detta dokument syftar till att på en övergripande nivå beskriva de olika komponenter och modelleringssteg som är involverade i processen bakom den virtuella panelen. Dokumentet kommer efter hand att kunna kompletteras med mer detaljerad information.



Historik

MMS lanserade AAM i maj 2021 för Program och klipp. Fram till och med 2022-09-25 användes en separat onlinepanel som ingående datakälla i modellen. Denna panel inkluderade endast mätning av digitala enheter (ej den stora skärmen) samt innehöll panelister 16 år och äldre. Detta innebar att modellen för AAM även inkluderade två specialsteg för att hantera de delar (dvs Big screen och barnmålgrupp) som inte mättes i panelen.

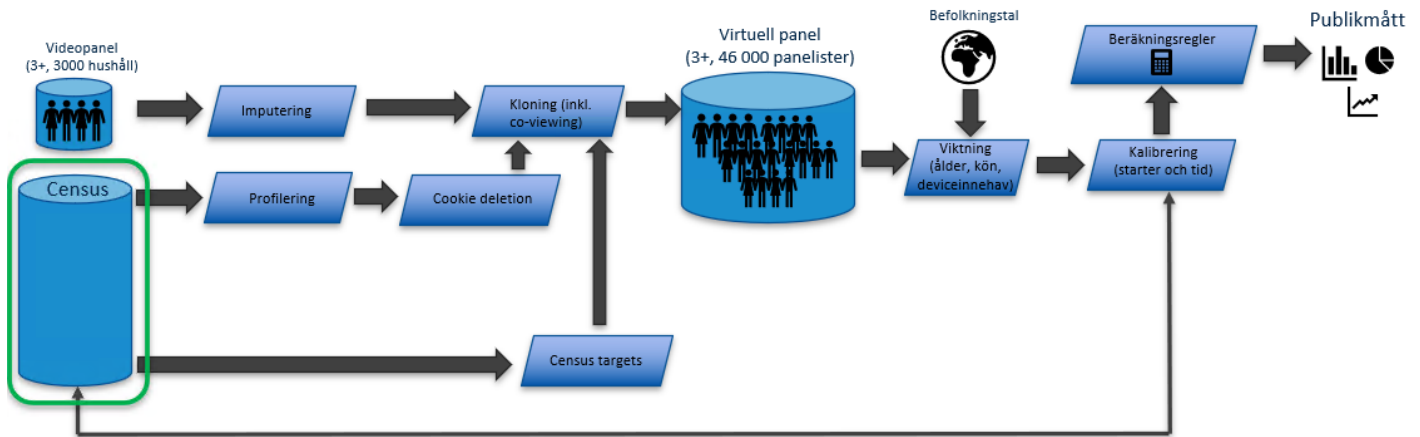
Från och med 2022-09-26 används i stället en single source videopanel där samtliga skärmar och åldrar (3+) mäts, vilket innebär att modellen från detta datum inte kräver några specialsteg utan alla skärmar och åldersgrupper kan modelleras på samma villkor.

En annan skillnad är att videopanelen inkluderar mätning av tillsammanstittande vilket medför att detta också kan inkluderas i den virtuella panelen, vilket ej var möjligt före 2022-09-26.

Datum	Händelse
2021-05-10	Lansering av AAM för Program & Klipp
2021-11-08	Övergång från vecko- till dagliga leveranser.
2022-09-26	Byte av leverantör för ingående datakällor (Census och panel) Exkludering av specialmodellering för Big screen och barnmålgrupp (detta mäts och modelleras nu på samma villkor som övriga delar). Inkludering av tillsammanstittande i AAM. Lansering av AAM för Reklam.

1 Datakällor

1.1 Censusdata

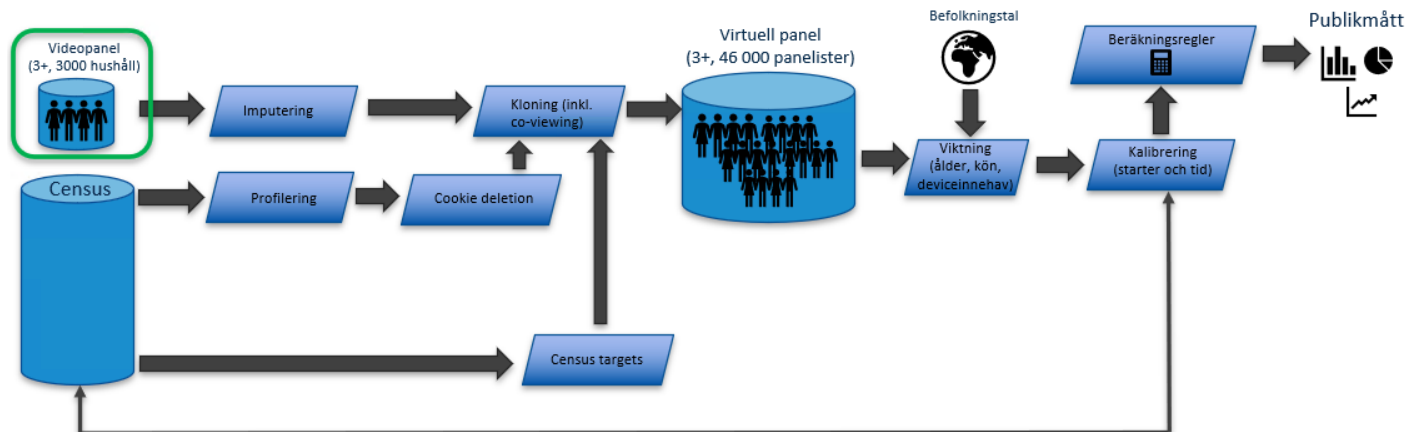


MMS mäter all reklamfilm och program/klipp som visats på de mätta videotjänsterna, där varje film/titel har en unik mättagg som fångas upp av MMS i alla de olika spelare och appar som mäts. Mätdata skickas till MMS som kvalitetskontrollerar, sammanställer och publicerar den officiella trafiken för både reklam och program/klipp. Censurmätningen möjliggör redovisning av totala volymer av starter och spenderad tid på titelnivå.

I censurmätningen fångas teknisk information om vilken slags enhet och browser som använts för videokonsumtionen, och varje videostart kan knytas till ett specifikt ID. Detta ID är i normalfallet en tredjeparts- eller förstapartscookie (vid konsumtion via browser) eller ett Ad-ID/IDFA (vid konsumtion via app), men det förekommer även andra typer av ID:n som så kallade "fallback-cookies" eller IDFV som används när det ordinarie ID:t inte kan användas (dvs när vare sig tredjeparts- eller förstapartscookie finns tillgängligt).

I varje given situation kommer det mest lämpade ID:t att användas för att i stor utsträckning som möjligt kunna knyta samman all trafik som kommer ifrån en och samma enhet. I resten av denna dokumentation kommer dessa id:n att benämnas "census-ID".

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 sammankoppling med ett census-ID.

MMS videopanel består av 3000 installerade hushåll (ca 6500 individer), där samtliga hushållets medlemmar i åldrarna 3-99 år utgör panelister. Videopanelen levereras av Nielsen, och är balanserad och viktad för att representera hela Sveriges befolkning och deras TV- och videokonsumtion. Inkluderat i videopanelen finns en virtuell subpanel på mellan 100-200 hushåll som saknar TV-apparat, till vilka liknande TV-hushåll donerar sin digitala konsumtion.

Konsumtionen mäts dagligen med hjälp av två olika mätare, People Meter (TV-tittande) och Streaming meter (Streamad video). Streaming metern möjliggör sammankoppling mellan panelisternas enheter och censusedmätningen. Detta gäller såväl digitala enheter som streaming devices kopplade till TV-skärmen, och tillåter därmed att panelisternas trafikdata från census kan identifieras.

Panelmätning inkluderar mätning av tillsammanstittande på TV-skärm, dator och surfplatta. Det gör att i paneldata kan samma uppmätta trafik från census kan tilldelas flera olika panelister, och denna data utgör sedan grunden för hantering av tillsammanstittande i AAM.

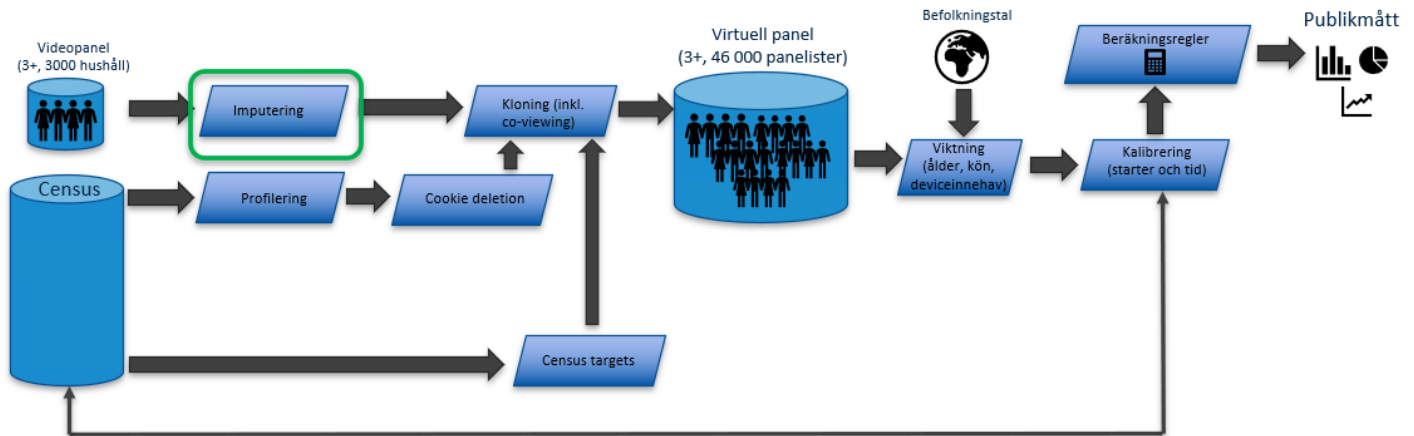
Att rakt av använda en panel och vikta upp dess uppmätta onlinevideokonsumtion är problematiskt på grund av den stora fragmentiseringen online. Många objekt (titlar/reklamfilmer) med låg räckvidd i befolkningen kräver en väldigt stor panelstorlek för att kunna mäta mer än endast de allra största titlarna/reklamfilmerna.

En stor panelstorlek medför dock andra svårigheter som att rekrytera tillräckligt många panelister och upprätthålla representativitet avseende både sociodemografi och mätta enheter.

Därför är det nödvändigt med kompletterande metodik för att kunna utnyttja den styrka som en kvalitativ panel utgör, men samtidigt möjliggöra redovisning av resultat på en betydligt mer heltäckande och detaljerad nivå.

2 Modellering

2.1 Imputering



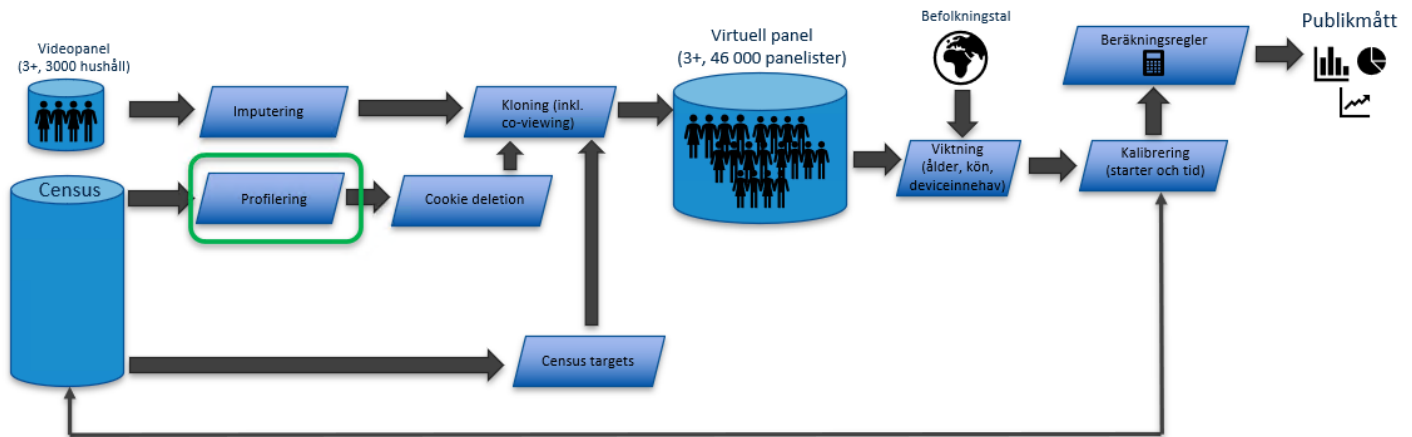
Även om det ställs höga krav på mätbarheten hos panelisterna och deras enheter så kommer det alltid förekomma fall där panelister/enheter inte är 100% mätbara, där definitionen av mätbar likställs med att enhetens trafik går att identifiera i censurmätningen. Det kan exempelvis handla om tekniska problem med hushållets streaming meter.

Tekniska svårigheter likt dessa kommer alltid finnas och medför vissa luckor i mätningen av panelisterna. Dessa behöver kompenseras för att inte leda till skevhet och underskattningar i resultatet. Detta sker genom så kallad imputering, vilket är en vanlig metod för att ersätta saknade värden i olika former av mätningar.

Imputeringen innebär i korthet att mätbara panelister agerar så kallade *givare* som "donerar" sitt tittarmönster till liknande panelister med luckor i sin mätning, så kallade *mottagare*. Givande och mottagande panelister paras ihop för att likna varandra med avseende på sociodemografisk profil, enhetstillgång och tittarmönster. En givare kan potentiellt donera sin konsumtion till flera olika mottagande panelister.

Målet med imputeringen är att ha en komplett panelmätning över samtliga skärmar som kärna för de fortsatta modellstegen.

2.2 Profilering



En grundbult i skapandet av den virtuella panelen är profileringsmodellen.

Profileringen syftar till att estimeras en sociodemografisk profil på varje enskilt census-ID, och möjliggör redovisning av antal starter och summa tid fördelat på olika målgrupper (åldersgrupp och kön). Den är dock inte tillräcklig för att på egen hand redovisa räckviddsmått i form av unika individer.

Det finns flera kända faktorer som gör att det inte går att likställa unika census-ID:n med individer, exempelvis:

- Ett ID kan tillhöra flera olika individer (vid delade enheter).
- En individ kan ha flera olika ID:n (t.ex 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.
- En och samma enhet/browser kan i olika situationer sammankopplas med flera olika census-ID:n.

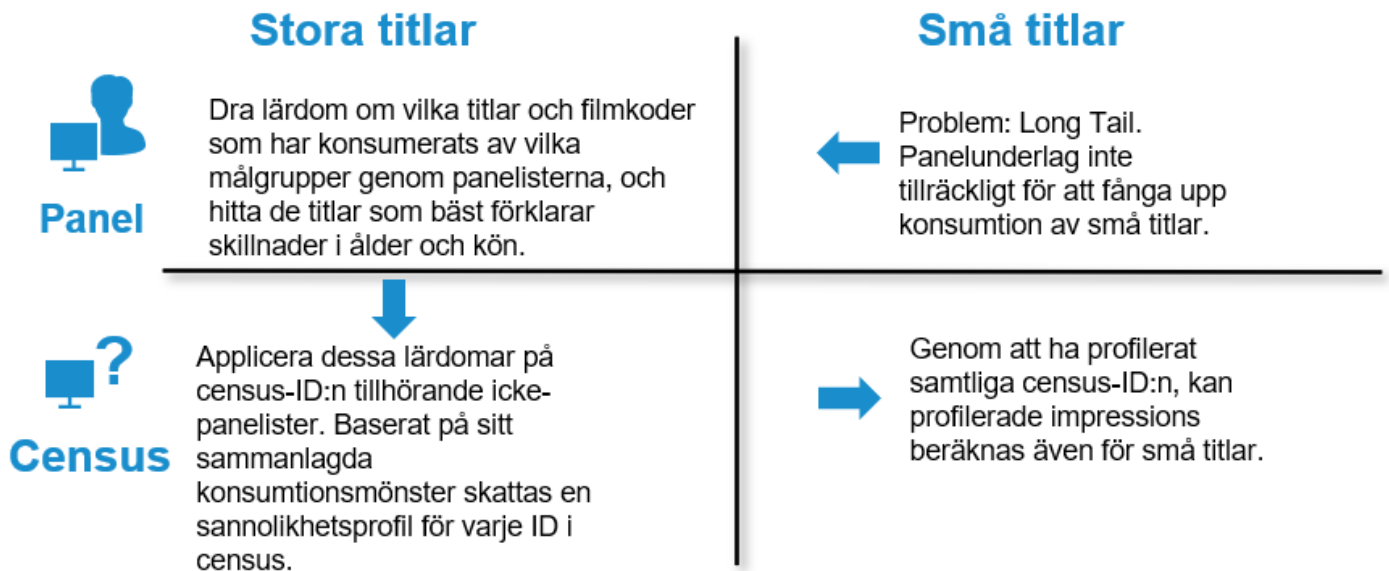
Därför krävs det alltså ytterligare hantering och modellering för att möjliggöra beräkning av räckviddsmått.

Målgruppsdata för Program och klipp baserades fram till och med 2021-05-09 helt och hållet på profileringsmodellen, men för data från och med 2021-05-10 då den virtuella panelen togs i drift så är profileringsmodellen alltså en av flera komponenter i den totala modellen.

Principen för profileringen kan beskrivas genom 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 titlar (long tail). Däremot kan vi för de tillräckligt stora titlarna/filmkoderna dra lärdomar av vilka målgrupper som konsumerat dessa i panelen.

Denna kunskap används för att skapa en *prediktionsmodell* som sedan appliceras på census-ID:n tillhörande icke-panelister, och tack vare att vi profilerar samtliga ID:n i census kan vi överföra kunskapen även på små titlar/kampanjer genom vetskapen om målgruppsprofilerna på de ID:n som konsumerat de små titlarna.



Varje dag används löpande den sammanlagda panelkonsumtionen från de senaste 28 dagarna. Denna datamängd används för att hitta samband mellan programtitlar och de olika målgrupperna. För varje kombination av titel och målgrupp skattas en korrelationskoefficient, där ett högt värde innebär ett starkt observerat samband mellan den aktuella titeln och målgruppen. De titlar med starkast samband, dvs bäst förklarar skillnader mellan kön och åldersgrupper, väljs ut att ingå i prediktionsmodellen.

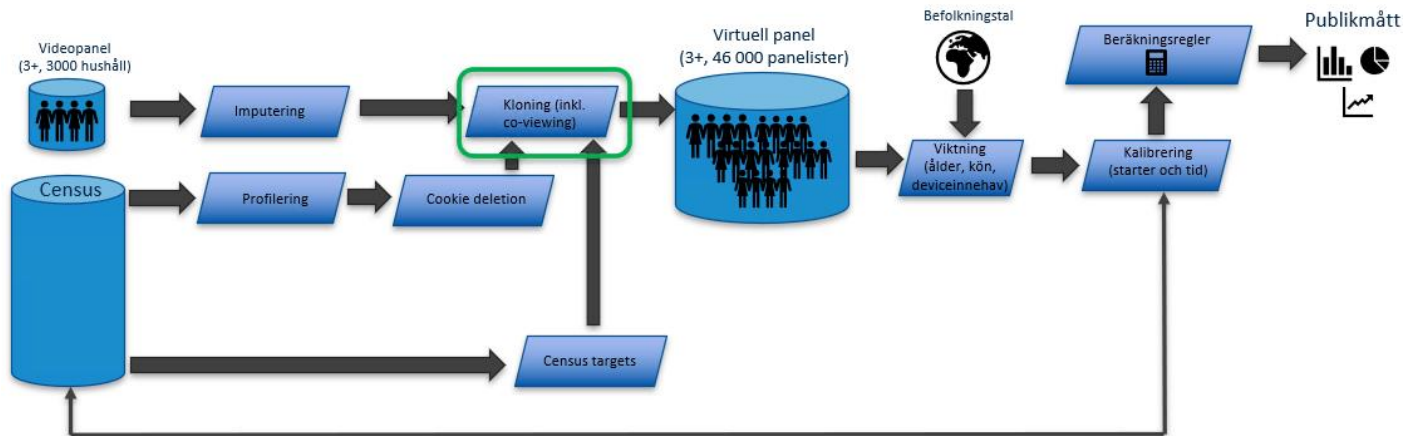
Prediktionsmodellen används sedan till att skatta demografiska profiler på alla ID:n i censusdata baserat på den konsumtion som observerats för respektive ID. Precis som i föregående steg används den sammanlagda trafiken från de 28 senaste dagarna för varje census-ID.

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

Profileringsmodellen kan aldrig 100% säkerhet att säga att ett ID tillhör en viss profil, och om det enbart har konsumerat innehåll som i panelen visat sig attrahera en väldigt bred publik så kan det vara svårt att exempelvis 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, där summan av alla målgrupper summerar till 1 (=100%).

2.3 Kloning



Med en rent panelbaserad mätning hade varje panelist i online-panelen tilldelats en vikt (=antalet personer i befolkningen som denne ska representera) och som använts för att vikta upp panelistens starter samt för att beräkna räckviddsmått.

Som tidigare nämnt är dock problemet att befintlig panelstorlek endast skulle räcka till att fånga upp en liten andel av allt innehåll i census, eftersom det finns så många titlar med liten konsumtion.

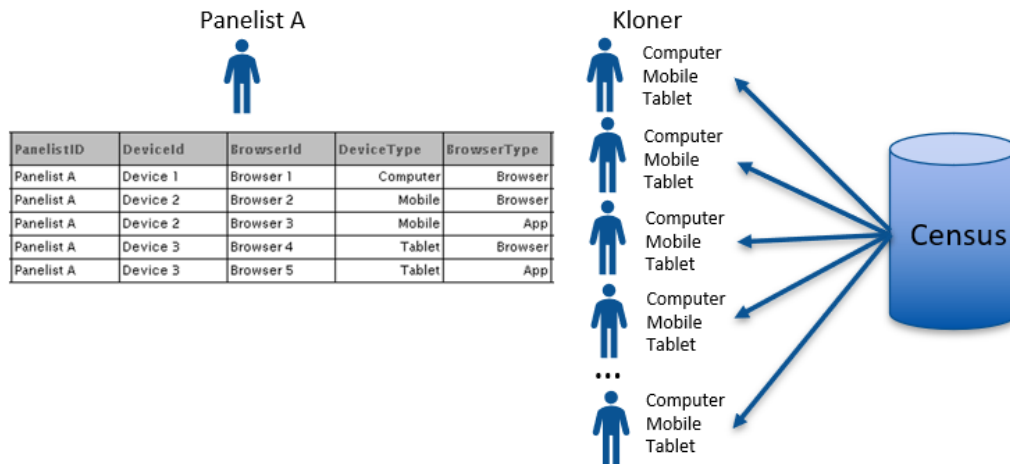
Det huvudsakliga syftet med den virtuella panelen är att utöka panelstorleken till en betydligt större panel, och detta görs genom att skapa kloner, så kallade virtuella panelister, med utgångspunkt från videopanelen.

Kloningen sker genom att virtuella panelister knyts till census-ID:n med beteendemönster som liknar originalpanelistens konsumtion. Faktorer som används i kloningen är exempelvis frekvens och intensitet (hur ofta och hur länge tittar man?), typ av innehåll (vilka sidor och typ av titlar tittar man på?) och sociodemografisk profil.

Det sistnämnda gör att profileringsmodellen beskriven i 2.3 är en grundsten i kloningen, eftersom virtuella panelister enbart kommer att kunna kopplas till census-ID:n som profilerats med en profil liknande den panelist som ska klonas.

En virtuell panelist kan vara sammankopplad med flera olika census-ID:n, på samma sätt som originalpanelisterna kan vara det eftersom de kan konsumera via flera olika typer av enheter och webbläsare.

De virtuella panelisterna kommer alltid att matcha sina respektive originalpanelisters innehav av olika typer av enheter. Det innebär att om en panelist har tillgång till exempelvis både smartphone, surfplatta och PC (Panelist A) i exemplet nedan, så kommer även samtliga dess kloner att ha tillgång samma kombination av enheter.



Varje enhet hos varje klon knyts till ett census-ID som matchar originalpanelistens beteendemönster på just den specifika enhetstypen. Om panelist ytterst sällan konsumerar någon av de mätta tjänsterna via PC, så kommer även dess kloner att ha ett väldigt lågfrekvent tittarbeteende på denna plattform.

Som en input i kloningsprocessen används trafikvolymerna från census per varje mätt site och plattform, samt aktivitetsnivåer i de olika målgrupperna. Dessa targetnivåer beräknas löpande för de 1,7 respektive 28 senaste dagarna, och den virtuella panelen för en specifik period designas för att i största möjliga utsträckning matcha targetvärdena på samtliga nivåer.

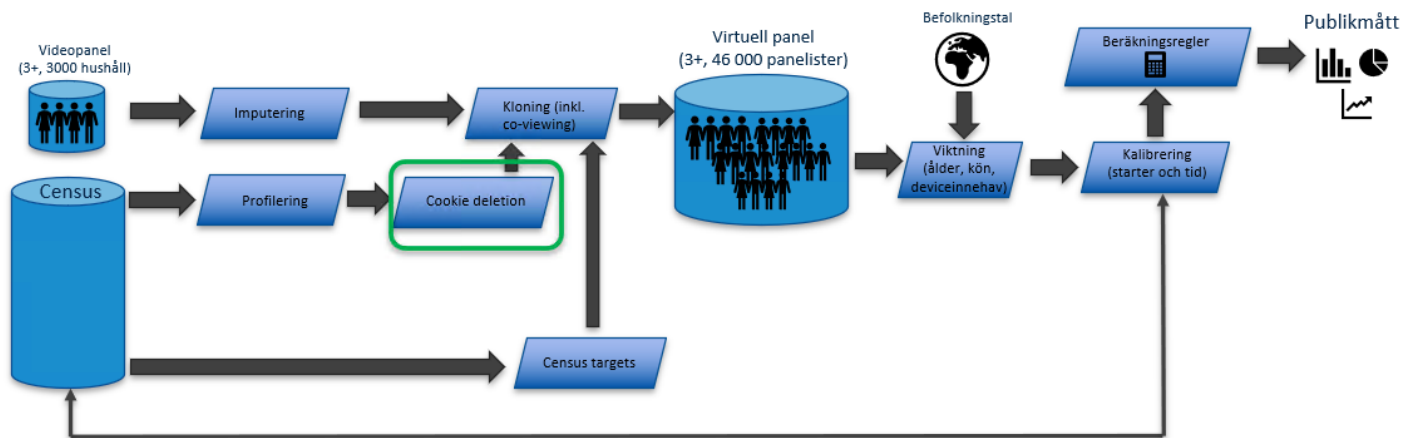
Ett verktyg för att uppnå detta är att tillåta variation i kloningsfaktorerna, dvs antal gånger som en specifik panelist klonas. Det innebär att antalet kloner som skapas av en specifik originalpanelist kan variera över tid. Det finns dock restriktioner på hur stor denna variation tillåts vara eftersom vi annars potentiellt skulle kunna riskera en instabil paneluppsättning med allt för stor omsättning över tid.

Designen av den virtuella panelen kan alltså beskrivas som en balansgång mellan att matcha resultatet från censusmätningen så väl som möjligt, men samtidigt bibehålla stabiliteten i panelen över tid.

I kloningsprocessen tas också särskild hänsyn till tillsammanstittande, för att detta tillskott av tittande ska kunna reflekteras i de slutliga tittarnivåerna. Videopanelens uppmätta tillsammanstittande ger lärdomar om tittarmönster gällande exempelvis sociodemografi, typ av innehåll, och när på dygnet tillsammanstittande främst sker. Denna data ligger sedan till grund för att skapa ytterligare kloner (och därmed tillföra ökad aktivitet) där det är motiverat utifrån den input som videopanelens tillsammanstittande ger.

Hanteringen av tillsammanstittande innebär inte att ett och samma tittarpass kommer att tilldelas flera samtidiga kloner, utan i stället syftar modellen alltså till att tillföra tillskott i tittandet via fler kloner. Detta innebär att det tyvärr inte är möjligt att utskilja vad som är singel- respektive tillsammanstittande i det slutliga resultatet, men att tillskottet från tillsammanstittande alltid är inkluderat i de nivåer som rapporteras från den virtuella panelen.

2.4 Cookie deletion-modell



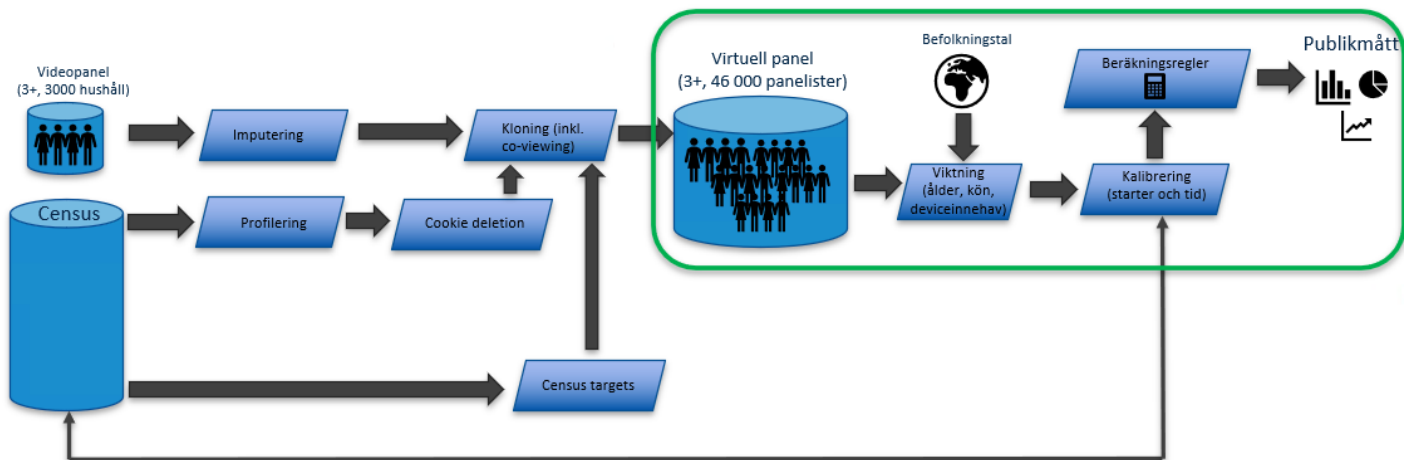
Som tidigare konstaterat kan cookies och andra ID:n i census förändras över tid.

Detta måste vi hantera genom att regelbundet substituera de ID:n som används för att sammankopplas med virtuell panelist. Utan detta skulle vi successivt tappa tittande i den virtuella panelen och få ett snedvridet och underskattat slutresultat.

Cookie deletion-modellen använder sig av en längre period av historiskt data (90 dagar) för att utifrån varje enskilt census-ID estimerar sannolikheten för att ett ID kan antas vara "död" och därmed behöver bytas ut med ett nytt med liknande egenskaper.

Om ett ID som historiskt setts i censusmätningen varannan dag plötsligt inte har någon observerad trafik under senaste veckan kan det alltså vara mycket sannolikt att detta ID är "död" och därmed måste ersättas. Medan ett ID med ett mer lågfrekvent tittande (exempelvis en gång i månaden) är mer sannolikt att fortfarande vara aktiv även om inte den heller har någon observerad trafik under senaste veckan.

3 Den virtuella panelen



3.1 Panelsammansättning

Den virtuella panelen kommer varje dag bestå av ca 46 000 rapporterande panelister i åldrarna 3-99 år. Varje klon, dvs virtuell panelist, kommer utgöra en specifik individ med ett eget panelist-ID i panelen. Initialt kommer den virtuella panelen att skapas veckovis, där panelens hela tittande för en komplett vecka ingår. Det innebär att paneluppsättningen kommer att vara densamma varje dag inom en kalendervecka, och att det kommer ske en viss fördröjning i resultatrapportering.

Nästa steg i utvecklingen är dock att möjliggöra dagliga leveranser av data, vilket kommer innebära en daglig variation i paneluppsättningen mer likt en traditionell panel (t.ex MMS TV-panel).

För varje virtuell panelist finns information om kön samt åldersgrupp vilket tillåter analys av alla publikmått på dessa variabler. De tillgängliga åldersgrupperingarna är konstruerade för att kunna kombineras ihop till de olika målgrupper som varit högst prioriterade bland MMS kunder. MMS ambition är att i framtida utveckling av modellen kunna bryta ner de större åldersgrupperna på 5-årsintervall.

3.2 Viktning och universumtal

Varje panelist har en daglig vikt som anger hur många individer i befolkningen som just denna panelist representerar. Panelisterna viktas utifrån kön, åldersgrupp och deviceinnehav. Det innebär att panelister med tillgång till exempelvis smartphone och PC viktas för att representera individer som är användare av just denna kombination av enheter, tillhörande samma kön och åldersgrupp.

Populationstalen som panelen viktas efter är beräknade utifrån MMS Basundersökning där vi på individnivå för alla åldrar 3-99 frågar om internetanvändning på respektive typ av plattform. Populationstalen kommer att uppdateras två gånger per år, i samband med TV-mätningens universumtalsskifte.



Nedan tabell visar exempel på hur populationen fördelade sig per olika cross device-kombinationer enligt Basundersökningen (2021:1):

Platforms	Befolkningstal (1000-tal)	% av Befolkningen
Ingen	411	4,1%
Big screen	261	2,6%
Big screenComputer	113	1,1%
Big screenComputerMobile	2831	28,4%
Big screenComputerMobileTablet	3067	30,7%
Big screenComputerTablet	84	0,8%
Big screenMobile	591	5,9%
Big screenMobileTablet	848	8,5%
Big screenTablet	272	2,7%
Computer	212	2,1%
ComputerMobile	538	5,4%
ComputerMobileTablet	362	3,6%
ComputerTablet	49	0,5%
Mobile	150	1,5%
MobileTablet	123	1,2%
Tablet	62	0,6%
	9974	100,0%

Observera att "Big screen" här inte är att likställa med tillgång till TV, utan ska tolkas som tillgång till att kunna streama till sin TV, exempelvis via Chromecast eller Apple-TV. Det är alltså möjligt att ha tillgång till TV utan "Big screen" (eller annan digital plattform).



Vid beräkning av relativa räckviddsmått (dvs % av befolkningen) så använder MMS hela målgruppens storlek i Sveriges befolkning, och dessa är samma som för motsvarande målgrupper i TV-mätningen. Befolkningstalen för målgruppsmodulerna i den virtuella panelen var enligt Basundersökningen 2021:1 följande:

Målgruppsmodul	Befolkningstal
K3-14	734611
K15-19	275879
K20-24	258326
K25-29	359235
K30-44	963865
K45-49	327769
K50-59	659355
K60-64	276827
K65-99	1105617
M3-14	752920
M15-19	286134
M20-24	329374
M25-29	380463
M30-44	1033888
M45-49	339324
M50-59	638511
M60-64	249458
M65-99	1001853
	9973409

Aktuella befolkningsstorlekar för varje universumtalsperiod från och med perioden 2021:1 kommer att tillgängliggöras på mms.se, för att kunna användas vid beräkningar i externa analysverktyg.

3.3 Kalibrering mot census

Som beskrivet i avsnitt 2.5 ingår en kalibrering mot censusnivåer i processen för att designa den virtuella paneluppsättningen. Denna kalibrering sker dock på en mer övergripande nivå, vilket innebär att den viktade panelens resultat nedbrutet på enskilda titlar inte behöver stämma exakt överens med antal uppmätta starter för samma titel i census.

För att ytterligare harmonisera med census sker därför en slutlig kalibrering mot de uppmätta nivåerna av starter och uppspelad tid i census. Denna kalibrering sker per kombination av dag, Site och titelkod. Detta säkerställer att det totala antalet starter för en specifik titel en viss dag är detsamma beräknat ur den virtuella panelen som i censusmätningen.

Observera dock att enbart titlar som har fångats upp av den virtuella panelen kan kalibreras. Det krävs alltså att en titel setts av åtminstone en virtuell panelist för att den ska kunna få ett resultat. De dagar titeln har fångats upp av panelen kommer antalet starter att överensstämma med censusmätningen, men sett



över längre perioder kan det alltså förekomma diskrepanser på grund av dagar där titeln inte alls setts av någon panelist.

Observera även att kalibrerade trafikmått per definition inte inkluderar tillsammanstittande, eftersom kalibreringen syftar till att uppnå överrensstämmelse med census och därmed inte täcker eventuellt tillskott via tillsammanstittande.

Den virtuella panelen har tack vare sin storlek en väldigt hög täckningsgrad och fångar upp i genomsnitt ca 90% av den totala trafikvolymen per dag. Av de titlar som inte fångas upp under en dag har ca 95% av dessa 100 eller färre starter under den aktuella dagen, och därmed väldigt låg sannolikhet att ha setts av någon i panelen. För titlar med mer än 5000 starter på en dag har vi hittills observerat 100% täckningsgrad, men det skulle teoretiskt kunna förekomma enstaka fall där även större titlar inte setts av någon panelist.

3.4 Beräkning av publikmått

För data från och med 2021-05-10 är samtliga publikmått för Program och klipp beräknade utifrån den virtuella panelen, och inkluderar därmed även beräkning av räckvidd och frekvens. Detsamma gäller Reklam från och med 2022-09-26.

Måtten beräknas i stor utsträckning enligt liknande metod som används för TV-valutan. Det innebär bland annat att vid räckvidd över en period av flera dagar så baseras beräkningen från en så kallad *paneldag*, där man utgår från paneluppsättningen och dess vikter just denna dag. Räckvidd över hela perioden beräknas sedan utifrån just denna paneluppsättning.

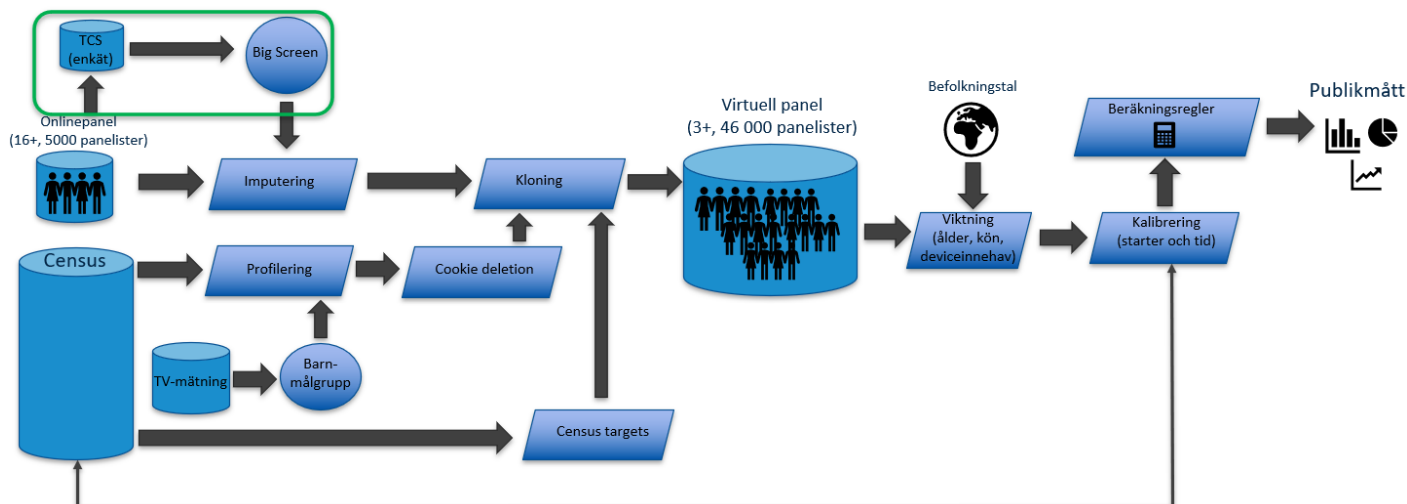
Då paneluppsättningen inte är identiskt varje dag så kan resultatet förändras beroende på vilken paneldag som används. Den officiella rekommendationen är att använda sig av periodens *första* dag.

För detaljerad beskrivning av hur de olika måtten beräknas, och hur AAM-data ska hanteras i externa system, hänvisas till MMS gyllene regler för beräkning [MMS gyllene regler för att beräkna mått från den virtuella panelen](#).

4 Appendix

Detta avsnitt inkluderar beskrivning av de särskilda modellsteg för Big screen och barnmålgrupp som krävdes innan övergången till single source videopanel. Dessa steg är alltså inte längre aktuella, men ligger till grund för resultatet fram till och med 2022-09-25.

4.1 Big Screen



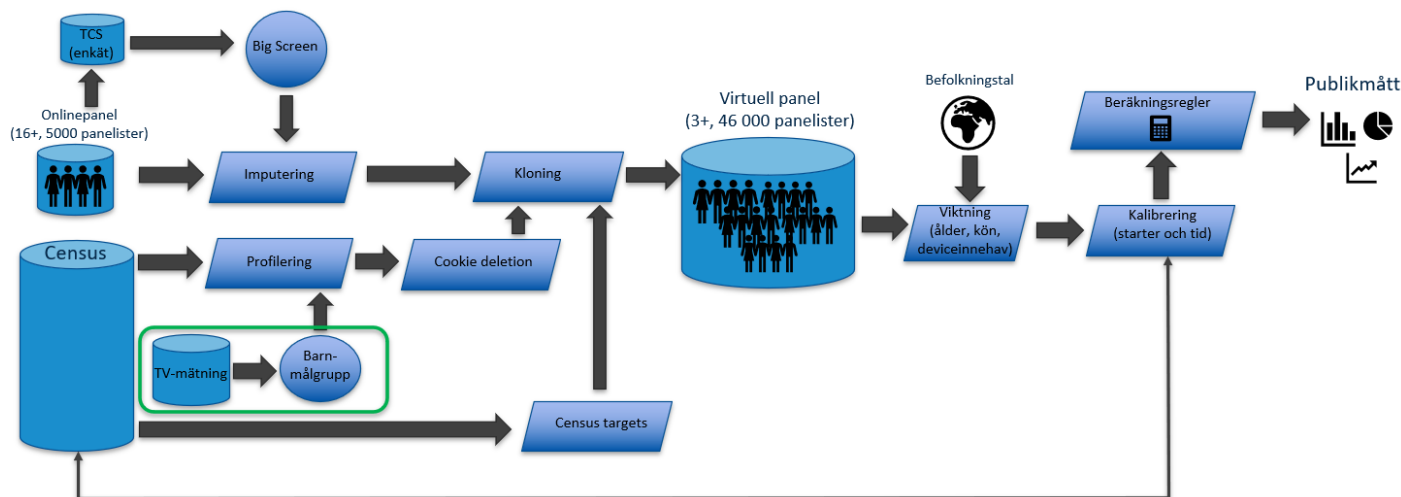
Imputeringen i 2.1 avser att fylla igen luckor av saknad mätning på smartphone, tablet och PC. När det gäller den stora skärmen/Big screen (dvs streaming via exempelvis Chromecast eller Apple-TV) saknas det i dagsläget möjlighet att koppla samman panelisternas streamingenheter med dess trafik i census.

Detta medför att samtliga panelister kan betraktas ha luckor i sin Big screen-mätning, och vi har alltså inga mätbara panelister som kan agera givare och donera sin konsumtion så som i det ordinarie imputeringssteget. Här behöver vi i stället ta hjälp av ytterligare en datakälla.

I onlinepanelen utförs en särskild enkät (också kallad "TCS") där panelisterna får svara på frågor om sin TV- och streamingkonsumtion på den stora skärmen. Enkäten innehåller frågor om frekvens (normalt antal dagar per vecka), uppskattad tittartid och typ av innehåll (genres) per respektive streamingtjänst.

Svaren från undersökningen ligger sedan till grund för att imputera Big Screen-konsumtion hos panelisterna. Panelister kopplas samman med census-ID:n vars trafik liknar det tittarmönster som panelisten uppgett i sina enkätsvar. I likhet med den ordinarie imputeringen i 2.1 kan man se census-ID:n som givare som donerar sin konsumtion till de panelister som enligt undersökningen brukar streama på den stora skärmen.

4.2 Barnmålgrupp



I dagsläget mäts inte barns tittande på online video i någon panel, vilket gör att vi saknar data för att lära oss om gruppens konsumtionsmönster.

För att trots detta kunna möjliggöra redovisning av denna målgrupp i resultatet har MMS skapat en specialmodell för just denna åldersgrupp.

Som input till modellering används målgruppens tittande på TV, uppmätt via MMS TV-panel som för närvarande är den källa där barnens tittande mäts på detaljerad (innehålls-) nivå.

I nedanstående bild visualiseras problemet på liknande sätt som tidigare, fast uppdelad på online-specifika titlar respektive titlar som visats både på TV och online via streamingtjänst:

	Titlar- TV+Online	Titlar - Online Only
<15 Online Panel	Problem: Mäts inte	Problem: Mäts inte
<15 TV Panel	Dra lärdom om hur målgruppen konsumerar olika titlar på TV, och överföra detta till samma titlar online enligt antagande om att titlar som lockar målgruppen på TV också gör det online.	Problem: Mäts inte
Census	Applicera dessa lärdomar på ID:n i census. Baserat på sitt konsumtionsmönster skattas en sannolikhet att tillhöra barnmålgrupp för varje enskilt ID..	Genom att varje census-ID har en skattad sannolikhet för ålder 3-14, kan profilerade impressions för <15 beräknas även för online-specifika titlar.

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 ett 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 census-ID kan skatta sannolikheten även att tillhöra "barnmålgruppen". Principen för detta kan förenklat beskrivas som att om ett census-ID 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 ID:t tillhör denna målgrupp.

För resultat fram till och med 2021-05-09 användas barnprofilerna, precis som ordinarie skattade målgruppsprofiler, för att redovisa antal starter fördelat per respektive målgrupp. I modellen för den virtuella panelen utgör dock profilerade census-ID:n basen för vilka ID:n som i nästa skede kommer kunna användas till *virtuella panelister*.

När barnmålgruppens tittande mäts (på innehållsnivå) även över digitala plattformar så kommer målgruppen att kunna hanteras på samma villkor som övriga målgrupper, och vi kommer därefter inte behöva detta specialspår som bygger på ett relativt grovt antagande.

4.3 Kloning av barnmålgrupp

Då barnmålgruppen inte mäts i online-panelen finns inga originalpanelister att skapa kloner utifrån på samma sätt som för övriga målgrupper.

En bas av artificiella barnpanelister behöver därmed skapas, med utgångspunkt i TV-panelens barnpanelister och deras konsumtion av innehåll som överlappar mellan TV och online (dvs TV-program som kan kopplas samman med en titelkod i onlinemätningen). Varje barnpanelist tilldelas slumpmässigt en kombination av enheter (exempelvis tablet + PC) så att den totala devicefördelningen matchar fördelningen i populationen för denna målgrupp baserat på MMS Basundersökning.

Varje census-ID profileras enligt avsnitt 4.2 med en sannolikhet att tillhöra barnmålgruppen. Därefter klassificeras alla ID:n att vara möjliga "barn-ID:n" eller inte baserat på om barnsannolikheten överstiger ett visst gränsvärde. Detta gränsvärde anpassas dynamiskt för att upprätthålla en stabil andel barnpanelister i den virtuella panelen över tid, där målsättningen är att fånga upp de ID:n som är mest sannolika att tillhöra barn och använda dessa för att koppla samman med virtuella barnpanelister.