MMS Golden Rules for calculating metrics for Total video

2024-09-05

Contents

0. Updates

Paragraph	Title	Date
	Document created	2024-01-31
4.1,4.2,4.4	Updated viewing measurements which use universe as a denominator in the calculation to use each day's universe. Also fixed some mistakes in the notation of overlap calculations.	2024-08-20

1. Introduction

From viewing date 2023-12-11 is MMS using the same model for reporting total video metrics for both ads and programs as well as on a total site/channel level.

This model creates connections between specific panelists in the AAM virtual panel and TV-panelists from the MMS video panel in such a way that online viewing from the AAM panel and TV viewing can be combined on a panelist level.

The model allows for calculations of total video metrics in a similar way as is being done in each of the two underlying currencies. This means that any external tool with access to the needed data sources can perform the same calculations as is being done in MMS reporting tool 'Totalanalys'.

The purpose of this document is to describe the necessary data sources, how they should be combined and the specific rules for calculating each of the metrics that can be calculated from this model.

The document covers only those metrics that are possible to calculate on a total video level. For details on how to calculate metrics for online or TV exclusively, please view the following documents:

MMS Golden Rules TV

MMS Golden Rules calculating metrics for online video

This document does not cover the methodological aspects behind the total video data. For information about MMS total video model, please view MMS Metodbeskrivning Totalvideo.

2. Data sources and definitions

2.1 Data files

To be able to calculate the metrics described in this document, the following data deliveries are needed from MMS:

Data file	Description
Panelistfil_total	File containing all daily combinations of AAM and TV panelists, and their daily weights to be used in the calculations.
Tittarfil_total_site	File containing all relevant viewing statements for both Online and TV viewing. To be used for reporting on a site/channel level.
Tittarfil_total_ad	File containing all relevant viewing statements for both Online and TV viewing on filmcode level.
Tittarfil_total_program	File containing all relevant viewing statements for both Online and TV viewing on program title level.
Metadata_total_ad	File containing metadata for the filmcodes included in tittarfil_total_ad
Metadata_total_program	File containing metadata for the program titles included in tittarfil_total_program
AAM_Universe_sizes	File containing population numbers for the sociodemographic groups available for total video metrics.

These files contain all relevant information required to report total video metrics according to MMS golden rules. No integration with TV and AAM specific data sources are needed.

The structure of the total video data sources is described in the following sections. Full data specifications can be found here.

2.2 Panelist data

Panelistfil_total is an extended version of the panelist file used in the AAM data delivery, where the AAM panelists are connected to panelists from TV data.

The combinations are being made with a many-to-many relationship. A **PanelistID** can be connected to several different **PanelistID_TV**, as is the case with **PanelistID**=21151 in the example below. The sum of weights for these combinations will equal the AAM weight for this **PanelistID** on this particular panel day.

	Date	PanelistID	PanelistID_TV	Age	Weight	Guest
1	20240101	NA	50831_17	A30-44	1821.80000	1
2	20240101	21151	377468_3	A15-29	160.17407	0
3	20240101	21151	963492_3	A15-29	5.24444	0
4	20240101	21151	134258_3	A15-29	26.32691	0
5	20240101	2683		A3-14	200.15462	0

Sum of weights equal AAM weight for PanelistID 21151

Similarly can a **PanelistID_TV** be connected to multiple different **PanelistID**, and the sum of weights for those combinations will equal the TV weight from [ORPI RP3 file](#).

As seen in the above example, there can also be **PanelistID**'s not connecting to any **PanelistID_TV**, or the other way around. These types of combinations represent online only users (or TV only users).

Each specific combination of **PanelistID*PanelistID_TV** has their own specific **Weight** and should in terms of reach calculations be treated as one unique panelist.

The **PanelistID*PanelistID_TV** combinations are unique per panel day and can vary across days.

2.3 Viewing data

The viewing data is delivered in three files, used for different reporting purposes (Ad, Program or total Site/Channel level). Each file contains both online and TV viewing sessions for the AAM and TV panelists.

When combining panelist and viewing data, all viewing sessions for a specific **PanelistID** should be allocated to every **PanelistID*PanelistID_TV** combination based on that panelist. The same applies to the TV viewing sessions belonging to a specific **PanelistID_TV**.

PanelistID	PanelistID_TV	Age	Weight	Guest
NA	50831_17	A30-44	1821.80000	1
21151	377468_3	A15-29	160.17407	0
21151	963492_3	A15-29	5.24444	0
21151	134258_3	A15-29	26.32691	0
2683		A3-14	200.15462	0

SiteID	Kanal	content_key	Date	PanelistID	PanelistID_TV	Source
1001029	NA	370971	20240101	21151		ONLINE
1001029	NA	366220	20240101	21151		ONLINE
1001029	NA	370676	20240101	21151		ONLINE
1001029	NA	362032	20240101	21151		ONLINE
1001029	NA	366222	20240101	21151		ONLINE
1001029	NA	366216	20240101	21151		ONLINE
1001029	NA	366216	20240101	21151		ONLINE
1001029	29	368511	20240101	NA	134258_3	TV
1001029	29	369747	20240101	NA	134258_3	TV
1001029	29	368976	20240101	NA	134258_3	TV
1001029	29	366966	20240101	NA	134258_3	TV
1001029	29	369782	20240101	NA	134258_3	TV
1001029	29	366848	20240101	NA	134258_3	TV
1001029	29	369346	20240101	NA	134258_3	TV
1001029	29	369453	20240101	NA	134258_3	TV
1001029	29	368725	20240101	NA	134258_3	TV
1001028	28	369522	20240101	NA	134258_3	TV
1001029	29	367022	20240101	NA	134258_3	TV
1001029	29	367160	20240101	NA	134258_3	TV
1001028	74	369012	20240101	NA	377468_3	TV
1001028	74	369621	20240101	NA	377468_3	TV
1001028	74	369621	20240101	NA	377468_3	TV
1001028	74	369012	20240101	NA	377468_3	TV
1001028	74	368576	20240101	NA	377468_3	TV
1001028	74	368576	20240101	NA	377468_3	TV

2.4 Viewing definitions

TV viewing is defined differently in the total video data compared to the TV currency. The TV currency is centered around broadcasts, where timeshifted viewing is allocated to the time of the live broadcast.

This is not an applicable perspective for total video, and instead all viewing sessions (for both TV and Online) will be allocated to the time when the actual viewing took place. That means that the column **Date** in the viewing files refers to the *viewing day*, and that all metrics derived from the total video model is defined as the viewing taking place during a selected period.

Another difference is that the TV currency includes all TV screen viewing of a broadcasted content (for up to 7 days), regardless of the viewing source. This means that part of the TV currency is streaming that is also measured as online big screen viewing within AAM.

That duplicated measurement will be excluded from the TV viewing in total video, and solely be included as online viewing measured through the AAM virtual panel.

This means that the TV viewing metrics derived from the total video model is *not* equivalent to the TV currency. Total video metrics should always be compared to the TV metrics calculated from total video data according to the rules in this document, and not to the TV currency definitions.

2.5 Objects of measurements

The smallest object of measurement is the **TitleCode/Filmcode** which is a unique ID referring to a specific program episode or to a specific ad. Calculations can be made over any aggregation of multiple titlecodes (or filmcodes), to get the total viewing on that group of programs during a period of time.

Reporting can also be done on a complete SiteID/Channel level, by using the **tittarfil_total_site** file. This file covers more viewing than the program and ad files combined, since it also includes for example TV viewing from channels that doesn't provide metadata.

Calculations can be made per viewing day, or across a period of several viewing days.

All metrics can be calculated and reported as either total cross-platform metrics, or per each **Source** (TV/Online) separately. As described in section 2.4, the TV viewing is defined differently compared to the TV currency.

2.6 Target modules

The total video data is currently limited to **Age** as the only sociodemographic information. There are five different age groups available, which are aggregates of the age groups available in AAM data.

2.7 Universe targets

Universe targets are used for calculating relative audience, and as a parameter in reach across multiple viewing days (section 4.4).

The universe figures equal the number of individuals in Swedish population derived from the establishment survey (MMS Basundersökning). They will be published on mms.se and updated twice a year.

When calculating daily metrics (for example average viewing time per day), it is always the universe numbers valid at each specific viewing day that should be used. For metrics across a period of time (for example relative cumulative reach) the universe from the *first* day of the period should be used.

3. Standards and recommendations

3.1 Reach criteria

MMS current standard for reach criteria is:

Type of analysis	TV	Online
Program	1 minute	15 seconds or at least 50% of video
Ad	1 minute	5 seconds or at least 50% of video
Site	1 minute	15 seconds or at least 50% of a video

The column **standard_reach_criteria** in viewing files flags whether the viewing sessions fulfill the standard criterion or not.

Other criteria can be used based on **Duration** only, where it's possible to combine different conditions for TV and online viewing statements.

A reach criterion is fulfilled if the length of the viewing statement in AAM viewing files is longer than or equal to the set condition.

3.2 Cumulative reach

Cumulative reaches should not be reported for periods longer than 28 days. This rule applies for all types of reporting (Program/Ad/Site).

MMS total video model uses target parameters calculated for the previous 1,7,14 and 28 days.

3.3 Limitations of small objects

Titlecodes with a total video reach less than 5000 individuals (A3-99) should not be included in any total video reporting.

This rule applies only for reporting at Program level.

4. Calculation of total video metrics

4.1 Rating/TRP

Rating is a measure of the number of non-unique contacts across the length of program, or across a complete day.

The base for rating is the accumulator of total viewing time calculated as:

$$Total\ viewing\ time\ = \sum_{v=1}^{V} (du_v * w_p * vf_v))$$

Where:
V= Total number of viewing statements ***v***.
du= Viewing time **Duration** for viewing statement ***v***.
w= Sum of **Weight** for all panelist combinations ***p*** belonging to viewing statement ***v***.
vf*=Viewing_factor** for viewing statement ***v.

Site

Rating on site/channel level is defined as the average number of viewers per second across the measured period. For the selection of **SiteID/Kanal**, the total viewing time is summarized across the complete period of ***D*** viewing days. and then divided by the sum of length ***L*** of all viewing days ***d*** in seconds:

$$Rating_{\ (d1+d2+...+dD)} = \frac{\sum_{d=1}^{D} Total\ viewing\ time_d}{\sum_{d=1}^{D} L_d}$$

Program

Rating on program level needs to be calculated separately for each **content_id (c)** before being aggregated into the rating for the selected **TitleCode (t).**

It is calculated as the average number of viewers per second across the **Length (L)** of the specific **content_id** according to **metadatafil_program:**

$$Rating_{\ (d1+d2+...+dD),c} = \frac{\sum_{d=1}^{D} Total\ viewing\ time_{d,c}}{L_c}$$

The rating for one or more **TitleCode (t)** is then calculated as the sum of the ratings for all individual content_ids:

$$Rating_{\ (d1+d2+...+dD),t} = \sum_{c=1}^{C} Rating_{d,c}$$

The rating can also be displayed in relative figures, where the absolute rating is divided by the universe size ***U*** of the selected target modules ***M*** for every day d, and then summed across all days of the selected period**:**

$$Relative\ rating_{\ (d1+d2+...+dD),t} = \sum_{d=1}^{D} \frac{Rating_{\ d,t}}{U_{M,d}} * 100$$

Ad

Rating (also called TRP) for ads needs to be calculated separately for each **Source** before being aggregated into to total video TRP.

Rating for **Source=**ONLINE is for a single **FilmCode (t)** calculated similarly as for a specific content_id for programs, where the total amount of viewing time is divided by the **Length (L)** of the **FilmCode** according to **metadatafil_ad**.

$$TRP\ Online_{\ (d1+d2+...+dD),t} = \frac{\sum_{d=1}^{D} Total\ viewing\ time_{d,t}}{L_t}$$

Rating for **Source=**TV is calculated in a similar way, but with the difference that the sum of total viewing time is always divided by the length of 60 seconds, and not the **Length** specified in the ad metadata file. This is due to the fact that the TV measurement is measured on a minute level and not as granular as the online measurement.

$$TRP\ TV_{\ (d1+d2+...+dD),t} = \frac{\sum_{d=1}^{D} Total\ viewing\ time_{d,t}}{60}$$

The total cross-platform rating is then calculated by adding the separate ratings as:

$$TRP\ Total_{\ (d1+d2+...+dD),t} = TRP\ Online_{\ (d1+d2+...+dD),t} + TRP\ TV_{\ (d1+d2+...+dD),t}$$

The rating across a number of ***T*** multiple filmcodes, for example a campaign, is calculated as the sum of individual rating for all included filmcodes:

$$TRP_{\ T} = \sum_{t=1}^{T} TRP_t$$

4.2 Average viewing time in population

Average viewing time is the number of minutes viewed in average per individual in the target population.

As for rating, the base is the accumulator of total viewing time calculated as described in section 4.1.

$$Total\ viewing\ time_{\ d} = \sum_{v=1}^{V} (d_v * w_p * vf_v))$$

Where:
V= Total number of viewing statements ***v***.
du= Viewing time **Duration** for viewing statement ***v***.
w= Sum of **Weight** for all panelist combinations ***p*** belonging to viewing statement ***v***.
vf*=Viewing_factor** for viewing statement ***v.

The average viewing time is then calculated by dividing the total viewing time with the universe target U of the selected target modules **M** for day *d* and then summarized across all days of the selected period:

$$Average\ viewing\ time_{(d1+d2+\cdots+dD)} = \sum_{d=1}^{D} \frac{Total\ viewing\ time_d}{U_{M,d}}$$

The formula above calculates the viewing time in seconds. When reporting average viewing-

time, the result should be converted to number of minutes.

The calculations are applied in the same way for a single **Source**, or on the complete dataset for total average viewing time across platforms.

4.3 Reach within a single viewing day

The rules described within this section should be applied when calculating the number of unique individuals viewing each day. The calculations can be performed for a single filmcode/titlecode/site, or across a selected group of multiple objects.

The calculations are applied in the same way for a single **Source**, or on the complete dataset for total cross-platform reach.

The formula below shows the calculation across a selection of titlecodes (Program), but is done in the same way for a selection of filmcodes (Ad) or sites/channels (Site).

For a titlecode (*t*) or a group of selected titlecodes (*T*) the absolute reach per each viewing-day (*d*) are to be calculated as:

$$Absoulte\ reach\ (single\ day)_{d,(t1,t2...tT)} = \sum_{p=1}^{P} w_p * D_{p,v}$$

$$D_{p,v} = \begin{cases} 1, & \text{if first matching } v \text{ for panelist } p \\ 0, & \text{else} \end{cases}$$

Where:
P= Total number of **PanelistID*PanelistID_TV** combinations in selected target-module(-s).
w= Weight for panelist combination ***p.***
v= Viewing statement for panelist combination ***p*** that matches the selection criteria.
D= A 'dummy'-variable that indicates if the viewing statement ***v*** should be considered in the reach calculation or not. Each individual panelist combination ***p*** should only be counted once, regardless the number of times the panelist has viewed any titlecode in the selection.

The accumulated sum of weights of unique panelist combinations could then to be divided with the universe size of the current target-module(-s*)* to calculate the relative reach:

$$Relative\ reach\ (single\ day)_{M,d,T} = \frac{Absolute\ reach\ (single\ day)_{M,d,T}}{U_M} * 100$$

Where:

U= The aggregation of universe sizes for all relevant target-modules ***M***.

Reach criterion

The reach calculation can be combined with a reach criterion stating that the panelist needs to have been viewing for at least X continuous seconds/minutes to be included.

The reach criterion should be based on the column **standard_reach_criteria** when using the standard criterion as described in section 3.1, or the column **Duration** for any other condition.

For the absolute reach formula in 4.3 this means that the 'dummy'-varible ***D*** should be set to 1 only for the first viewing statement of each panelist combination ***p*** that meets the selected condition.

Calculation across multiple objects

Reaches for separate objects cannot be aggregated with each other. The reach calculation should

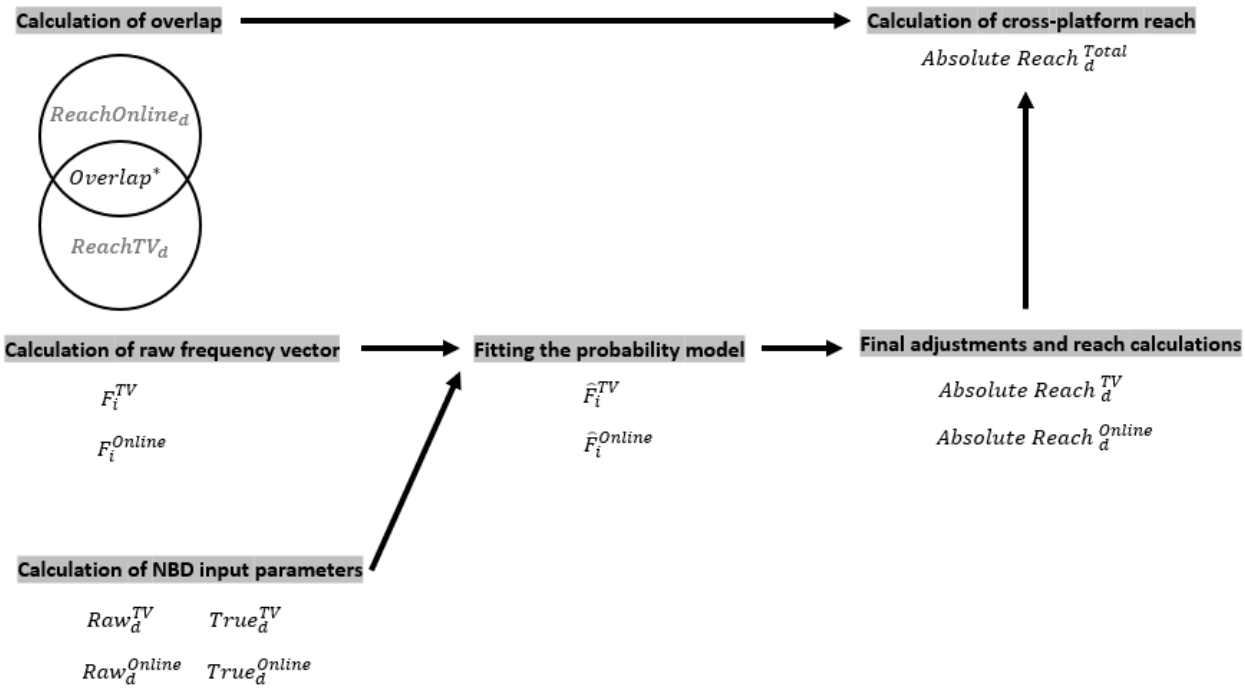
always be performed directly on the complete selection of unique titlecodes/filmcodes etc.

4.4 Reach across multiple viewing days

This section refers to calculation of unique viewers over a period covering several different viewing days. The calculation can be performed on a selection of multiple titlecodes (or filmcodes or sites/channels) combined.

The calculation differs a bit depending on if the **Source** is TV or Online, and to calculate total cross-platform reach it is needed to always calculate both sources separately and apply a set of estimated overlap parameters as a final step.

This process visualized below is repeated in an iterative way by adding each day within the period one at time for d1, d1+d2, d1+d2+d3…etc up until the complete *n* days of the selected period.



Panel-day

In this kind of calculation, a certain *panel-day* should be used. For total video metrics, the panel_day should always be the *first* day of the selected period for calculating the reach levels for Ad and Content separately.

The base for the calculation is the set of all possible **PanelistID***PanelistID_TV panelist combination, including their weights, from the **Panelistfil_total** belonging to the panel-day. Only this set of panelist combinations and their viewing during the selected viewing-period are to be considered when calculating the number of unique viewers across the selected period. Any viewing made by panelists not belonging to any panelist combination at the panel-day should not be included in the calculation.

To correct for differences between the reporting samples for each specific day and the used sample belonging to the panel-day, MMS will use the negative binomial distribution (NBD) to model the final frequency distribution. This is the same method as used for Reach and Frequency in both the TV and online currencies.

Reach criterion

The calculation of reach across multiple days can be combined with a reach criterion in a similar way as for reach within a single day (4.3).
The calculation of the below described input parameters *True, Raw* and the frequency distribution F_i should then be limited to the set of viewing statements where the viewing fulfills the selected criteria.

Calculation of overlaps

This section is needed to calculate the total cross-platform reach. For calculation of reach per a specific **Source**, please skip this section and begin with "Calculation of raw frequency vector".

Please note that overlap estimation is performed on the raw reach levels before applying the NBD adjustment. That means that $Reach$ in this section is always referring to the reach levels calculated directly from the panel-day sample.

The overlap estimation is done in an iterative process over all viewing days in the selected period of n viewing days.

Beginning with the first day $d1$ we calculate the raw reaches per each **Source** separately and combined as:

$$(ReachOnline_{d1}|pd = d1) = \sum_{p=1}^{P} w_p * D_{p,v}$$

$$(ReachTV_{d1}|pd = d1) = \sum_{p=1}^{P} w_p * D_{p,v}$$

$$(ReachTotal_{d1}|pd = d1) = \sum_{p=1}^{P} w_p * D_{p,v}$$

$$D_{p,v} = \begin{cases} 1, & \text{if first matching } v \text{ for panelist } p \\ 0, & \text{else} \end{cases}$$

Where:
pd= Panel-day for the panel sample to be used for the calculation.
P= Total number of **PanelistID***PanelistID_TV combinations in selected target-module(-s).
w= Weight for panelist combination ***p.***
v= Viewing statement for panelist combination ***p*** that matches the **Source** selection criteria.
D= A 'dummy'-variable that indicates if the viewing statement ***v*** should be considered in the reach calculation or not. Each individual panelist combination ***p*** should only be counted once, regardless the number of times the panelist has viewed any titlecode in the selection.

From that we can derive the number of persons viewing both TV and Online for day $d1$ as:

$$(ReachBoth_{d1}|pd = d1) = \big((ReachTV_{d1}|pd = d1) + (ReachOnline_{d1}|pd = d1)\big) - (ReachTotal_{d1}|pd = d1)$$

For the first day $d1$ applies that:

$$ReachBoth^{*}_{d1} = (ReachBoth_{d1}|pd = d1)$$

For day $d2$ we continue to calculate the raw reaches per **Source** for days d1+d2 in total, still using the panel sample from $pd = d1$.

$$(ReachOnline_{d1+d2}|pd = d1)$$

$$(ReachTV_{d1+d2}|pd = d1)$$

In addition, we calculate the overlap for day $d2$ using the panel sample from $pd = d2$ as:

$$(Overlap_{d2}|pd = d2) = 1 - \frac{(ReachTotal_{d1+d2}|pd = d2) - (ReachTV_{d1+d2}|pd = d2)}{(ReachOnline_{d1+d2}|pd = d2)}$$

After that we calculate:

$$TargetReachBoth_{d2} = (Overlap_{d2}|pd = d2) * (ReachOnline_{d2}|pd = d1)$$

$$(\Delta ReachOnline_{d2}|pd = d1) = (ReachOnline_{d2}|pd = d1) - (ReachOnline_{d1}|pd = d1)$$

$$(\Delta ReachTV_{d2}|pd = d1) = (ReachTV_{d2}|pd = d1) - (ReachTV_{d1}|pd = d1)$$

There are now three restrictions that needs to be fulfilled in order to estimate $ReachBoth^*_{d2}$. These are:

$$ReachBoth^*_{d2} \geq ReachBoth^*_{d1}$$

$$ReachBoth^*_{d2} \leq (ReachTV_{d2}|pd = d1)$$

$$ReachBoth^*_{d2} \leq (\Delta ReachOnline_{d2}|pd = d1) + (\Delta ReachTV_{d2}|pd = d1) + ReachBoth^*_{d1}$$

Which consequently means that $ReachBoth^*_{d2}$ can be calculated as:

$$minReachBoth^*_{d2} = max(TargetReachBoth_{d2}, ReachBoth^*_{d1})$$

$$\mathrm{maxReachBoth}^*_{\mathrm{d2}} = \min\left(minReachBoth^*_{d2}, (ReachTV_{d2}|pd = d1)\right)$$

And finally:

$$ReachBoth^*_{d2} = \min(\mathrm{maxReachBoth}^*_{\mathrm{d2}}, (\Delta ReachOnline_{d2}|pd = d1) + (\Delta ReachTV_{d2}|pd = d1) + ReachBoth^*_{d1})$$

This process is repeated for each day dk up until day dn. Where we for each day calculate reach online, reach TV and the overlap as:

$$(ReachOnline_{d1,+d2+\cdots+dk}|pd = d1)$$

$$(ReachTV_{d1+d2+\cdots+dk}|pd = d1)$$

$$(Overlap_{dk}|pd = dk) = 1 - \frac{(ReachTotal_{d1+d2\ldots+dk}|pd = dk) - (ReachTV_{d1+d2+\cdots+dk}|pd = dk)}{(ReachOnline_{d1+\mathrm{d2}+\cdots+dk}|pd = dk)}$$

And uses them to calculate $ReachBoth^*_{dk}$ using the above-described steps where $_{d2}$ is substituted with $_{dk}$ and $_{d1}$ is substituted with $_{dk-1}$.

The condition $pd = d1$ remains fixed during all iterations, while $pd = d2$ is substituted with $pd = dk$

The final overlap estimates for day dn can then be calculated as:

$$Overlap^*_{dn} = \frac{ReachBoth^*_{dn}}{(ReachOnline_{dn}|pd = d1)}.$$

Which will be applied on the final NBD-adjusted **Source**-specific reach levels as described in section *Calculation of cross-platform reach*.

Calculation of raw frequency vector

For each **Source** separately, we calculate the "raw" frequency vector of the number of individuals who have viewed exactly 1 viewing day, exactly 2 viewing days…exactly n viewing days during the selected period.

For each **PanelistID*PanelistID_TV** combination belonging to the panel-day, all viewing statements (fulfilling the reach criterion) for all viewing-days in the selected period should be used to count the number of active viewing-days within the period. All viewing from guest viewers (**Guest=**1) should be excluded.

By summarizing the weights for the **PanelistID*PanelistID_TV** combinations who have viewed exactly 1 viewing-day, exactly 2 viewing-days…exactly n viewing-days as:

$$F_i = \sum_{p=1}^{P} pw_p$$

$$\text{for } i = 1,2,3,\ldots,n$$

Where:
P= Total number of possible **PanelistID*PanelistID_TV** combinations belonging to the panel-day, according to the panelist file.
pw= Weight for panelist combination ***p*** at the panel-day, according to the panelist file.
i= Number of days viewed within the period

The raw frequency distribution will be adjusted by fitting the negative binomial distribution to it as described in the following sections.

Calculation of input parameters

This step describes how to calculate the input parameters $True$ and Raw which are needed for the NBD modelling.

These parameters are calculated in slightly different ways for TV and Online, as it's being done in each of the separate currencies.

For **Source=**ONLINE the parameters are calculated as:

1. The $True$ number of contacts are basically the total sum of contacts for the selection of titlecodes (or filmcodes/Sites) and complete period of n viewing-days.

$$True_{(d1+d2+\ldots+dn)} = \sum_{v=1}^{V} (c_v * w_p * cf_v)$$

Where:
V= Total number of relevant viewing statements ***v*** in viewing file.
c= Number of uncalibrated **Contacts** for viewing statement ***v.***
w= Sum of **Weight** for all panelist combinations ***p*** belonging to viewing statement ***v.***
cf= **Impression_factor** for viewing statement ***v.***

2. The *raw* number of contacts are calculated in a similar way, but with one important difference. Only the viewing statements belonging to panelist combinations included in the reporting sample of the panel-day $pd = d1$ are to be considered. Also, the panelists weights from the panel-day should be used.

   The raw number of contacts for the selection of T titlecodes and D viewing-days are then calculated as:

$$Raw_{(d1+d2+...+dn)} = \sum_{v=1}^{V} (c_v * pw_p * cf_v)$$

   Where:
   V= Total number of relevant viewing statements ***v*** in viewing file.
   c= Number of uncalibrated **Contacts** for viewing statement ***v.***
   pw= Sum of **Weight** for all panelist combinations ***p*** belonging to viewing statement ***v.***
   cf*=** **Impression_factor** for viewing statement ***v.

For **Source=**TV the parameters are calculated as:

1. The $True$ rating is the total sum of rating across all days and C objects included in the selection. The rating is calculated according to the description in section 4.1.

$$True_{(d1+d2+...+dn)} = \sum_{c=1}^{C} Absolute\ Rating_{,c}$$

2. The Raw rating is calculated in a similar way, but only the viewing statements belonging to **PanelistID***PanelistID_TV combinations included in the reporting sample of the panel-day $pd = d1$ are to be considered, and guest viewers (**Guest=**1) are excluded. This means that the panelist combination weights from the panel-day should be used when calculating the total viewing time accumulator used for the rating.

$$Raw_{(d1+d2+...+dn)} = \sum_{c=1}^{C} Absolute\ raw\ rating_c$$

Fitting the probability model

The below steps are applied in the exact same way regardless of the **Source** in question:

1. Calculate the relative frequency distribution by dividing each element of the input frequency vector with the universe size of the selected target-module(-s):

$$Fr_i = \frac{F_i}{U_M}$$

$$for\ i = 1,2,3, \dots, n$$

2. Calculate the share of individuals watching exactly 0 days of the period by summarizing all elements of the frequency vector calculated in the previous step, and subtract this from 1:

$$Fr_0 = 1 - \sum_{i=1}^{n} Fr_i$$

   Special case: In case $\sum_{i=1}^{n} Fr_i > 1$, let element $Fr_0 = 0$.

3. Add element Fr_0 to the frequency vector Fr to make it cover $i = 0,1,2, \dots, n$ number of active viewing days.

   Special case: In case Fr begins with one or more elements where $Fr_i = 0$, these elements should be removed from the frequency vector before applying the probability model in the following steps.

   Example: Assume the following relative frequency distribution:
   Exactly 0 days: 0
   Exactly 1 day: 0
   Exactly 2 days: 0.5
   Exactly 3 days: 0.4
   Exactly 4 days: 0
   Exactly 5 days: 0.1

   Since the distribution begins with two elements with value 0, these two elements should be removed from the frequency vector. The remaining Fr_i will then cover the frequency counts:
   $i = 2,3,4,5$.

4. Before being able to apply the model, it is necessary to estimate the needed NBD parameters. The first step is to estimate the parameter *k:*

$$k = abs\left[-\frac{RawR * ln(Fr_0)}{ln(Fr_0) - RawR * W_{-1}\left[\frac{Fr_0^{\left(\frac{1}{RawR}\right)} * ln(Fr_0)}{RawR}\right]}\right]$$

Where:

RawR= The relative version of the input parameter *Raw*, where it is divided by the universe size of the relevant target-module(-s):

$$RawR = \frac{Raw}{U_M}$$

Fr_0 = The value of the 0-element in the frequency vector, as calculated in step 2.
W_{-1}[ ]= The Lambert W function, using the secondary branch.

<u>Special case:</u> In case $k > 120$, let $k = 120$.

5. After estimating the value of *k,* the parameters *a* and *b* can be calculated as:

$$a = \frac{\frac{Raw}{k}}{U_M}$$

Where:

U: The aggregation of universe sizes for all relevant target-modules M.

And:

$$b = a * \frac{True}{Raw}$$

6. With the values of *a* and *b,* the NBD probabilities of *p_raw* and *p_true* can be calculated as:

$$p\_raw = \frac{a}{(a+1)}$$

$$p\_true = \frac{b}{(b+1)}$$

7. With all needed parameters, each element of the frequency distribution Fr will be adjust by using the function of the negative binomial distribution:

$$\widehat{Fr}_i = \frac{\Gamma(k+i)}{\mathrm{i!}\,\Gamma(\mathrm{k})}[(p\_true)^i(1-p\_true)^k - (p\_raw)^i(1-p\_raw)^k] + Fr_i$$

Where:

Γ( )= Gamma function.
i= The number of $i = 0,1,2,\ldots,n$ active viewing days as remaining after step (3.).

Final adjustments and reach calculation

1. The next step is to adjust the sum of the frequency shares to make it equal to 1 (=100% of the population).

   In case the sum of $\widehat{Fr} \neq 1$, the difference should be added by adjusting the last, n^{th}, element of the frequency distribution as:

$$\widehat{Fr}_n = \widehat{Fr}_n + \left(1 - \sum_{i=0}^{n} \widehat{Fr}_i\right)$$

   <u>Special case:</u> In case any elements of the frequency distribution where removed due to the special case rule in step (3.), these should now be included again (with their values of 0) to make the full adjusted frequency distribution $\widehat{Fr}_i$ cover the complete number of $i = 0,1,2,\ldots,n$ number of active viewing days.

2. The relative reach across the selected period can now be calculated as the sum of elements $i = 1,2,\ldots,n$, i.e the share of individuals who have watched during at least one of the viewing-days within the period:

$$Relative\ reach\ (multiple\ days) = \sum_{i=1}^{n} \widehat{Fr}_i$$

Calculate the absolute frequency counts by factor all elements of the relative frequency distribution $\widehat{Fr}$ with the universe size of the relevant target-modules(-s):

$$\hat{F}_i = \widehat{Fr}_i * U_M$$

The sum of the absolute frequency distribution $\hat{F}$ gives the absolute number of reached individuals during the selected period:

$$Absolute\ reach\ (multiple\ days) = \sum_{i=1}^{n} \hat{F}_i$$

Cumulative reach over time

The reach calculation can be used to display the growth of additional reached individuals day by day during the selected viewing period.

All above steps including calculation of input parameters, fitting the probability model and final reach calculation needs to be done in an iterative way adding each day within the period one at a time for d1, d1+d2, d1+d2+d3 etc up until the complete *n* days of the selected period.

The panel-day, d1=20240101, is chosen based on the complete selected period and should be used for each specific iteration within the period.

	$\hat{F}_1$	$\hat{F}_2$	$\hat{F}_3$	$\hat{F}_4$	$\hat{F}_5$	***Absolute Reach+1***
d1=20240101	100 000					**100 000**
d2=20240102	80 000	30 000				**110 000**
d3=20240103	75 000	27 000	8 000			**110 000**
d4=20240104	70 000	32 000	4 000	4 000		**110 000**
d5=20240105	65 000	35 000	10 000	2 000	1 000	**113 000**

Calculation of cross-platform reach

After calculating the final cumulative reaches for both TV and Online, it is time to apply the overlap parameters estimated in section *Calculation of overlaps* to get the total cumulative reach across both **Sources.**

For each day dk up until *n* days in the selected period, the $Overlap^*_{dk}$ is applied to the corresponding cumulative absolute reach levels as:

$$Abs.Reach_{dk}^{Total} = Abs.Reach_{dk}^{TV} + Abs.Reach_{dk}^{Online} * (1 - Overlap^*_{dk} *)$$

The absolute reach levels can be divided with the universe figure ***U*** for the selected ***M*** target-modules to get the relative total reach:

$$Relative\ Reach_{dk}^{Total} = \frac{Absolute\ Reach_{dk}^{Total}}{U_M} *100$$