

UPU-Agreed Measurement Systems External Audit 2021

Universal Postal Union
International Bureau

Audit Report January 2022



1. Executive Summary

The UPU GMS has been running Quality of Service measurements since 2009, starting with 21 countries. This number rose over the years reaching 65 countries in 2021 that participated in the UPU Quality of Service link to terminal dues (commonly referred to as UPU QS link). Similarly, the International Postal Corporation (IPC) has been running the UNEX measurement system, for which ten countries were measured and participated in the QS link in 2021. Having two UPU-agreed measurement service providers (MSPs) measuring 75 countries for the purposes of the QS link calls for transparency and reliability in the measurement output to provide the confidence needed going forward in the quality of postal service delivery not only in each country measured but also at on a global level.

As a proven and reputable audit services provider, PwC was pleased to support UPU with this challenge, leveraging our extensive experience in the postal industry, particularly in quality monitoring and auditing.

In agreement with the UPU Directorate of Postal Operations (DOP), we performed audit activities for the two MSPs, GMS and UNEX, using the UPU Global Monitoring Technical Design 3rd Edition (UPU GMS TD) with the following scope:

- Reperformance on samples of calculation of statistical design and allocation of links and items for the year 2021
- Reperformance on samples January–June 2021 of item validation
- Verification of bundling on samples for the period January–June 2021
- Analysis of differences in measurements between the first six months of 2020 and 2021
- Follow-up on recommendations in the report 2020

2021 was expected to be the year of recovery from the devastating impact of COVID-19, but, as experienced by most of the countries, new waves of the virus and jeopardized logistics processes continued to deeply influence public health, society and economies around the world. Despite increased digitalisation and changes in behaviour that helped the world start the adaptation to the ‘new normal’, several challenges needed to be mastered.

Countries, including the ones measured by the Universal Postal Union Global Monitoring System (UPU GMS) were affected by restriction measures, health concerns that impact people who are part of the global measurement endeavour and by the still existing disruption of flows. A valuable effort has been made to keep the regular production of test letters and the panellist network alive and efficient, and with success, as can be seen from the available figures.

In our audit procedures, we have additionally analysed some key measurement figures as ‘allocation’ and ‘valid on target’, comparing the first six months of 2020 with the first six months of 2021. The figures are clearly indicating how the world was affected by waves of the pandemic, with different timings and amplitudes across regions. Based on the analysed data, we have gained a good impression of the performed measurements and the recovery pattern. Despite recovery, the disruption of the logistics flows is still visible in the high number of bundled items on the inbound side.

Reperformance on samples of calculation of statistical design and allocation of links and items for the year 2021

In close contact with key contact persons at the UPU International Bureau (IB) and IPC, we performed a recalculation of the statistical design and allocation of links and items for a sample of two countries of the UNEX UPU measurement (a level A and a level B) and three countries of the GMS UPU measurement (a level A, a level B and a level D), based on the rules of the UPU GMS TD and without finding any deviation.

Reperformance on samples January–June 2021 of item validation

All items for the five countries in the sample (the same as for the statistical design calculation reperformance) in the months of April, May and June 2021 were verified as being correctly marked as On-time or Delayed.

The correct application of validation rules as stated in Appendix G of the UPU GMS TD (P1.1, P1.2, P1.5, P1.7) has been verified. Minor exceptions have been noted only for rule P1.1 (timely registration of dropping). These exceptions are reported in the findings but have no impact on the measurement. For rule P1.4 (bundling), refer to the bundling paragraph.

Verification of bundling on samples for the period January–June 2021

The same sample as in the reperformance validation has been also used for the verification of bundling.

Bundling on the sending side has been observed, both related to allocation and to delayed dropping. The number of deviations is very limited, mostly related to single panellists. These deviations can be explained as a pandemic effect, with the panellists not being able or not feeling safe to drop off the item as planned. This bundling has no direct effect on the measurement but are increasing the risk that bundling occurs on the inbound side. We do not consider these deviations as a non-compliance, but we are suggesting improving the controls (refer to the finding paragraph).

Bundling on the inbound side has been generally high for the countries in the sample, in some cases reaching 25% of the items, exceeding the thresholds defined by the UPU GMS TD. This bundling is not generated by bundling on the sending side, which was minor, but from the disruption in logistic flows due to the pandemic, with items from some countries being bundled on a reduced number of days. Bundling requests from designated operators could affect the measurement. We do not consider this bundling a non-compliance, but we suggest the POC to address the possible issue arising in future from the application of a systematic bundling check as defined in paragraph 18.1.7 of the UPU GMS TD. To be noted that it is responsibility of the designed operator to request a bundling check and that in future, in line with the UPU GMS TD automatic bundling check could be implemented by the Measurement Service Providers (MSP).

Analysis of differences in measurements between the first six months of 2020 and 2021

Since two of the most visible consequences of the COVID-19 pandemic on measurement were the real allocation of items to links and the reduced number of Valid on Target items (VOT), we have been comparing data related to these topics for the first six months of 2020 and 2021. The aim was to understand if a recovery took place and if doubts around accuracy exist independently of the re-weighting procedure.

The key results of the analysis are the clear general improvement of the average VOT, being relatively stable over 70% from the beginning of the year, reaching almost 90% in May and June.

We also see that, for most of the countries, the first six months of 2021 were largely better than the first six months in 2020, even if a few countries, especially in Europe, where still struggling.

Findings of the previous year

The two findings (Finding ID 1–2: retention and training of panellists) from the 2018 audit, which only partially affect compliance and have low significance, are currently still open. These findings are related to conscious decisions made to improve operational processes that are not yet reflected in the current UPU GMS TD document.

We note that the Compliance and Audit Process (CAP) expert team has been working with IPC and China Post to identify the source of the problem reported in 2018 for ‘Test items not reaching destination’ (Finding ID 3 in 2018) and has been urging the UPU member countries already implementing or joining the UPU QS Link to make every effort to use passive technology. There is currently one country in the UNEX UPU TD measurement struggling with the implementation of passive technology for reasons going beyond the jurisdiction of the operator. Due to the difficulties to technically address the problem raised by travel restrictions related to the pandemic, no usable data for the year 2021 will be available for the performance measurement of this country.

General result

Based on our procedures as described in this report, nothing came to our attention that caused us to believe that the activities performed by UPU GMS, by UNEX UPU TD measurement systems or by the service providers in the audited areas were not compliant with the UPU GMS TD document.

This report has been prepared solely for the use of UPU in connection with the audit as requested by the UPU IB and should not be quoted in whole or in part without our prior written consent. No responsibility to any third party is accepted, as the report has not been prepared for and is not intended for any other purpose.

The procedures performed by us do not constitute either an audit or a review made in accordance with International Standards on Auditing or International Standards on Review Engagements. Consequently, we do not express any assurance on the information included in this report.

Table of Contents

1. Executive Summary	1
2. Scope of our work	4
3. Audit Methodology and Process	5
4. Audit results.....	7
Annexes.....	17
A1 Rating Criteria.....	18

2. Scope of our work

The main objective of the external audit was to assess whether the methodology, its implementation and the calculation of QS measurement results by the two MSPs were compliant with the UPU GMS TD document in 2021

The scope covered the following areas and components:

- Verification of the statistical design (incl. allocation of links and items)
- Recalculation of the Performance Measurement
- Recalculation of validation of item
- Verification of bundling
Calculation and reporting of Quality-of-Service results



3. Audit Methodology and Process

Based on our postal measurement experience, we developed specific audit procedures that we applied in this engagement.

We performed an assessment of the current postal measurement procedures that will allow UPU to understand the quality of service they are getting from their service providers in comparison with what is required by the UPU GMS TD 3rd Edition document. We also provide clear insight on where improvements are needed and clear enforceable recommendations.

Our approach is:

- Independent
- Comprehensive
- Reliable and robust
- Statistically accurate
- Quality-driven and standardised
- Tested and proven over many years
- ISO9001 consistent

While the methodology is standardised, PwC recognises that each client's environment and requirements are different. Hence, we customised it for this specific task, focussing on the four areas in respect of compliance to the UPU GMS TD document:

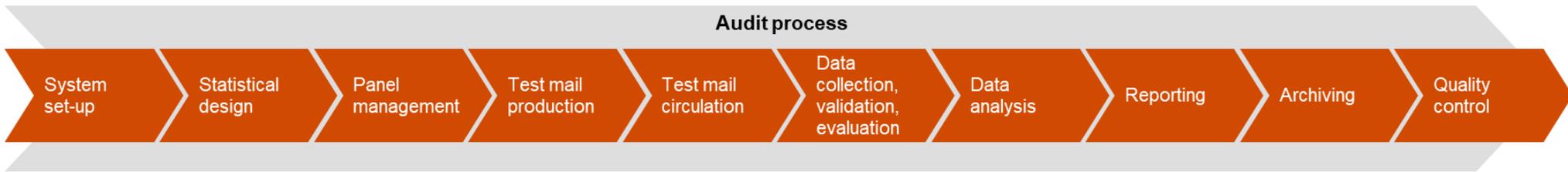
Calculation and reporting of Quality-of-Service results

- Panel Management
- Quality control and validation
- RFID Diagnostic Monitoring

The focus of the assessment of 2021 was set on quality of service (allocation, validation and performance calculation): Therefore, our methodology was this year underpinned by the following tasks: Understanding the requirements of the UPU GMS TD document.

- Assessing the risks and mapping all elements in focus to our specific audit process (ref. diagram 1). We produced a viable, solid and efficient work plan.
- Collect information in appropriate mode: we know what should exist and how it can be assessed.
- Obtaining during the UPU and IPC interviews information and documentation by exchanging experience on postal measurement management with like-minded PwC people.
- Understanding deviations and confirming them with follow-ups. Performing recalculations wherever appropriate, using the UPU GMS TD as basis for confirmation and leveraging our specific tools for this purpose.
- Formulating preliminary reports that can be validated.
- Producing a final report that is adequate for management and for those who have to work with it.
- Findings are formulated in a way that will help follow-up actions and improvements.

This methodology was used from the first year, confirming situation and progress, leveraging all the experience from previous years.



Process execution against design at MSP and at organisation managing the systems

<p>Operations and IT set-up</p> <p>Panel set-up and organisation</p> <p>Quality assurance set-up</p>	<p>Data collection validation, organisation and transmission for implementation of statistical design</p> <p>Application of GMS technical design, especially geographical requirements, in implemented statistical design</p>	<p>Recruitment of panellists</p> <p>Panel performance management through KPIs</p> <p>Incentive management</p> <p>Panel training</p>	<p>Generation and preparation of test mail items</p> <p>Programming and integration of RFID tags</p> <p>Dispatch of test mail items</p>	<p>Test mail circulation</p> <p>Registration of induction and delivery information and return of test mail items</p>	<p>Data entry and validation of panellist induction and delivery data</p> <p>Evaluation of panellist data</p> <p>Validation of panellist data against RFID data</p> <p>Diagnostic monitoring</p>	<p>Data analysis</p> <p>Exception reporting</p> <p>Proactive analysis to identify potential project risks</p> <p>Accuracy of calculations</p>	<p>KPI(s)</p> <p>Reporting according to timetable</p> <p>Recommendation from site survey process</p>	<p>Archiving of text mail items</p>	<p>Contingency planning</p> <p>Quality controls</p> <p>KPIs</p> <p>Change management process</p> <p>Process monitoring</p>
--	---	---	---	--	--	---	--	-------------------------------------	--

Demonstrating understanding of GMS technical design by Measurement Service Partner (MSP)

 <p>Existence and extent of documentation for all audited areas</p>	 <p>Correct application of GMS technical design</p>	 <p>Implementation of country specific design parameters</p>	 <p>Implemented internal controls framework</p>
--	--	---	--

4. Audit results

4.1. Results per audited area

Based on our procedures performed, nothing came to our attention that caused us to believe that the activities performed by UPU GMS, by UNEX UPU TD measurement system or by the service providers in the audited areas were not compliant with the UPU – GMS Technical Design 2nd Edition document.

The following table provides an overview of the results over the audited areas. When we noted at least one non-compliant finding, we marked the area as red; otherwise, it is marked yellow when there was at least one partially compliant finding. Areas are marked as green when no compliance issues were detected in the given area. The numbers included in the table below indicate how many findings were identified per measurement area (in total 2, see detailed list in chapter 4.2

Measurement Areas	UNEX UPU TD meas.	UNEX UPU TD meas. – PMC – Kantar	UPU GMS meas.
A. Statistical design (sample design)			
B. System configuration and inputs			
C. Panel management			
D. Mails production			
E. Mails circulation (distribution/sending/receiving)			
F. Data collection, validation and processing			
G. Transit time calculations			
H. Statistical Analysis			
I. Reporting			
J. Archiving			
K. Quality Control			
L. RFID Diagnostic Monitoring system			

Compliance rating: Compliant Partially compliant Non-compliant

4.2. Detailed findings

The following list shows the current identified and open findings.

Finding ID	Area ID	Area Description	Assessment Area	Compliance	Issue description	Significance	Recommendation / Assessment results
1	C1	Panellists' recruitment questionnaires to ensure that UPU-specific recruitment requirements are satisfied	UNEX UPU TD measurement - PMC - Kantar	 Partially compliant	<p>Panellists' retention period</p> <p>The panellists were not informed, as part of the hiring process, about the requirement that they should be willing to participate for at least six months. This is not fully in accordance with chapter 7.2 of the UPU GMS TD document: "In all cases, panellists: [...] should be willing to participate for at least six months".</p> <p>However, we noted that the approach generally used to reduce the risk of not having the necessary number of panellists is not addressed by formally requesting the panellist to commit for at least six months but by having and managing back-up panellists.</p> <p>Finding remains open and unchanged in 2021.</p>	 Low	<p>We recommend either implementing a clause in the recruitment questionnaire to ensure the panellist is aware that he/she is expected to participate for at least six months or agreeing with UPU on updating the formulation of the UPU GMS TD.</p> <p>The UNEX UPU TD measurement system and Kantar do not fully agree with the recommendation as they express concerns because being formally bound by such a retention requirement may put off panellists from staying at least six months on the panel.</p> <p>Therefore, we suggest the UPU GMS measurement system and UNEX UPU TD measurement system formally agree on the next steps and assess whether the recommendation needs to be implemented or the formulation of the UPU GMS TD can be adjusted.</p>

Finding ID	Area ID	Area Description	Assessment Area	Compliance	Issue description	Significance	Recommendation / Assessment results
2	C6	Process of panellists' training	UNEX UPU TD measurement - PMC – Kantar	 Partially Compliant	<p>Training of Panellists</p> <p>There was no formalised way to assess whether panellists have been sufficiently trained before starting to act as a panellist. However, we noted that the panellists' performance was monitored and that, in the case of low performance, the panellist was retrained.</p> <p>The UPU GMS TD document (chapter 7.3) mentions that "training should confirm that the panellist has understood the task involved and is able to carry it out as instructed".</p> <p>In addition, the documented training programme for newly recruited panellists does not cover the topics on how to indicate the condition of the item received (envelope damaged, address label damaged or not fully legible, transponder missing, etc.). This is not fully in line with UPU GMS TD document (chapter 7.3.2) where it states, "instructions should indicate: [...] how to indicate the condition of the item received (envelope damaged, address label damaged or not fully legible, transponder missing, etc.)".</p> <p>Finding remains open and unchanged in 2021.</p>	 Low	<p>We recommend implementing an assessment process to ensure the knowledge of the panellist is tested before involving her/him as an active panellist.</p> <p>In addition, we recommend adding to the instructions provided to panellists a section on how to indicate the condition of the item received.</p> <p>The UNEX UPU TD measurement system and Kantar do not fully agree with the recommendation as they express concerns because they believe that training guidelines (via video, long-form written and FAQs) provide a comprehensive introduction to panellists' tasks. In addition, they monitor their panellists to confirm that they understand their duties. If deviations are observed, panellists will be retrained or dropped as appropriate.</p> <p>Therefore, we suggest the UPU GMS measurement system and to UNEX UPU TD measurement system formally agree on the next steps and assess whether the recommendation needs to be implemented or the formulation of the UPU GMS TD can be adjusted.</p>

Finding ID	Area ID	Area Description	Assessment Area	Compliance	Issue description	Significance	Recommendation / Assessment results
3	-	-	UNEX UPU TD measurement – UPU GMS		<p>Test items not reaching destination</p> <p>We noted that the UNEX UPU TD measurement system was affected by a large number of test items not reaching destination, even after a long period of time, despite being induced according to the UPU GMS TD document.</p> <p>In particular, no items at all induced between June 2018 and October 2018 reached the destination countries. Starting in November 2018, items were registered again: 12 out of 1,178 in November 2018 and 33 out of 1,141 in December 2018. Please refer to section Update 2019 for the current situation.</p> <p>Since the items were produced in line with the UPU GMS TD document and there are no indications they were not induced, this is not considered as a non-compliance matter, but the number of valid test items falling below the recommendations of the UPU GMS TD is influencing the performance measurement of the receiving countries. No similar pattern for the UPU GMS has been identified. The issue is known to UNEX UPU TD measurement and to UPU GMS, but no root cause has yet been identified.</p> <p>Update 2021:</p> <p>We acknowledge that the CAP has been running a pilot with IPC and China Post aiming to identify the root of the problem and to resolve it. The pilot has been generating basic information on the problem, identifying issues related to the problem and formulating recommendations.</p> <p>The consequences of the problem have been addressed by introducing the re-</p>	<p>🟡 Medium</p>	<p>It is recommended to continue the investigation as planned by the CAP together with IPC and China Post until a clear resolution is found.</p> <p>It is also recommended to make all possible efforts to resolve the situation for the operators having external difficulties introducing passive technology.</p>

Finding ID	Area ID	Area Description	Assessment Area	Compliance	Issue description	Significance	Recommendation / Assessment results
					<p>weighting procedure. The investigation will be continued with a new pilot.</p> <p>The CAP has also urged the UPU member countries already implementing or joining the UPU QS Link to make every effort to use passive technology.</p> <p>There is currently one country in the UNEX UPU TD measurement struggling with the implementation of passive technology for reasons going beyond the jurisdiction of the operator. Due to the difficulties to technically address the problem raised by travel restrictions related to the pandemic, no usable data will be available for the performance measurement of this country.</p> <p>The resolution and the activities have been strongly affected by the COVID emergency.</p> <p>Finding remains open in 2021, medium significance remains unchanged.</p>		
4	F	Data collection, validation and processing requirements	UNEX UPU TD measurement – UPU GMS		<p>We observed for the sampled countries some limited bundling on allocation.</p> <p>We also noted some low bundling caused by a delayed dropping of some panellists.</p> <p>To be noted that this bundling on the sending side is contributing to the much larger bundling seen on the inbound side (in some cases reaching 25% of the items), possibly driven by the COVID related disruptions in the international logistics (refer to the point of attention regarding Cap in paragraph 4.4).</p>	🟡 Medium	<p>We suggest</p> <ol style="list-style-type: none"> 1) To further reduce the bundling in allocation. 2) To monitor bundling caused by deviations from droppers. <p>See also the point of attention regarding Cap in paragraph 4.4.</p>

Compliance rating:  Compliant  Partially compliant  Non-compliant

Significance rating:  Low  Medium  High

4.3. Analysis of differences in measurements between the first six months of 2020 and 2021

The COVID-19 pandemic has been impacting the world dramatically from a human perspective, with many lives lost and causing pain for many, and creating a state of emergency in the health, social and economic system. The virus has been affecting the world in several waves starting in the first quarter 2020. The time frames of the waves differed between regions and countries, with some period of apparent recovery in-between.

In 2020, the disruption of transportation and logistics affecting the measurements has been blocking flows, some of them completely. In some cases, even the production of test letters was suspended.

In 2021, an increased digitalisation of processes and availability of health support instruments (vaccines and better pandemic containment measures) reduced the lockdowns and the impact on logistics.

KPIs of the GMS measurement were largely affected. We have been analysing in particular how 'Allocation' and 'VOT' changed in the first six months of 2020 and 2021 respectively, across all measured countries.

Allocation

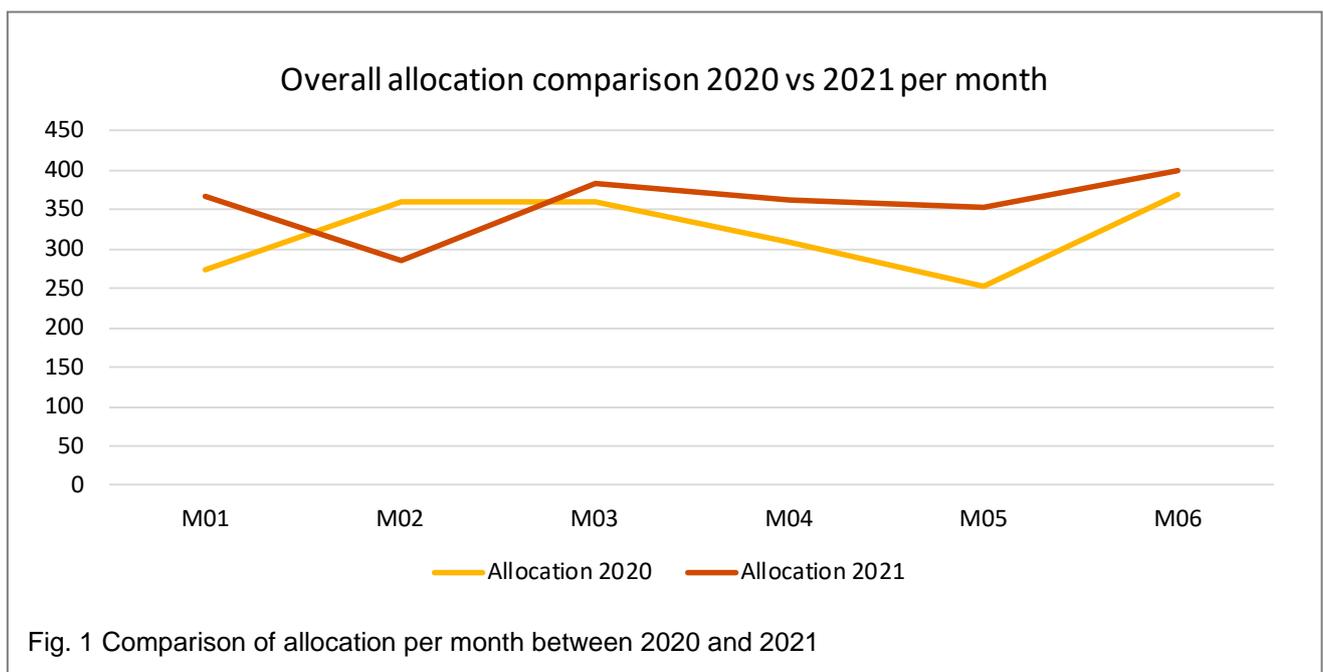


Fig. 1 Comparison of allocation per month between 2020 and 2021

In Fig. 1, the average monthly allocation is shown. It can be noted how the production of test items was steadily reduced in 2020, from February through May. Allocation ramped up again in June.

In 2021, the allocation following the pandemic was reduced again in February to raise anew in March and then remain relatively stable.

It can be noted that allocation was never shut down completely, apart from that one time at the UK production site in May 2020.

It has been important to keep the network of panellists alive and, in that sense, test items were continuously dispatched, preparing contingency plans to produce and send even in lockdown situations or to resume production after a shutdown has occurred.

Valid on Target

The impact of the pandemic is much more visible by looking at VOTs, since this indicator is affected by several factors, covering several aspects of the measurement chain, including lockdowns for panellists and employees of the postal services, ordered lockdowns and logistical jeopardy.

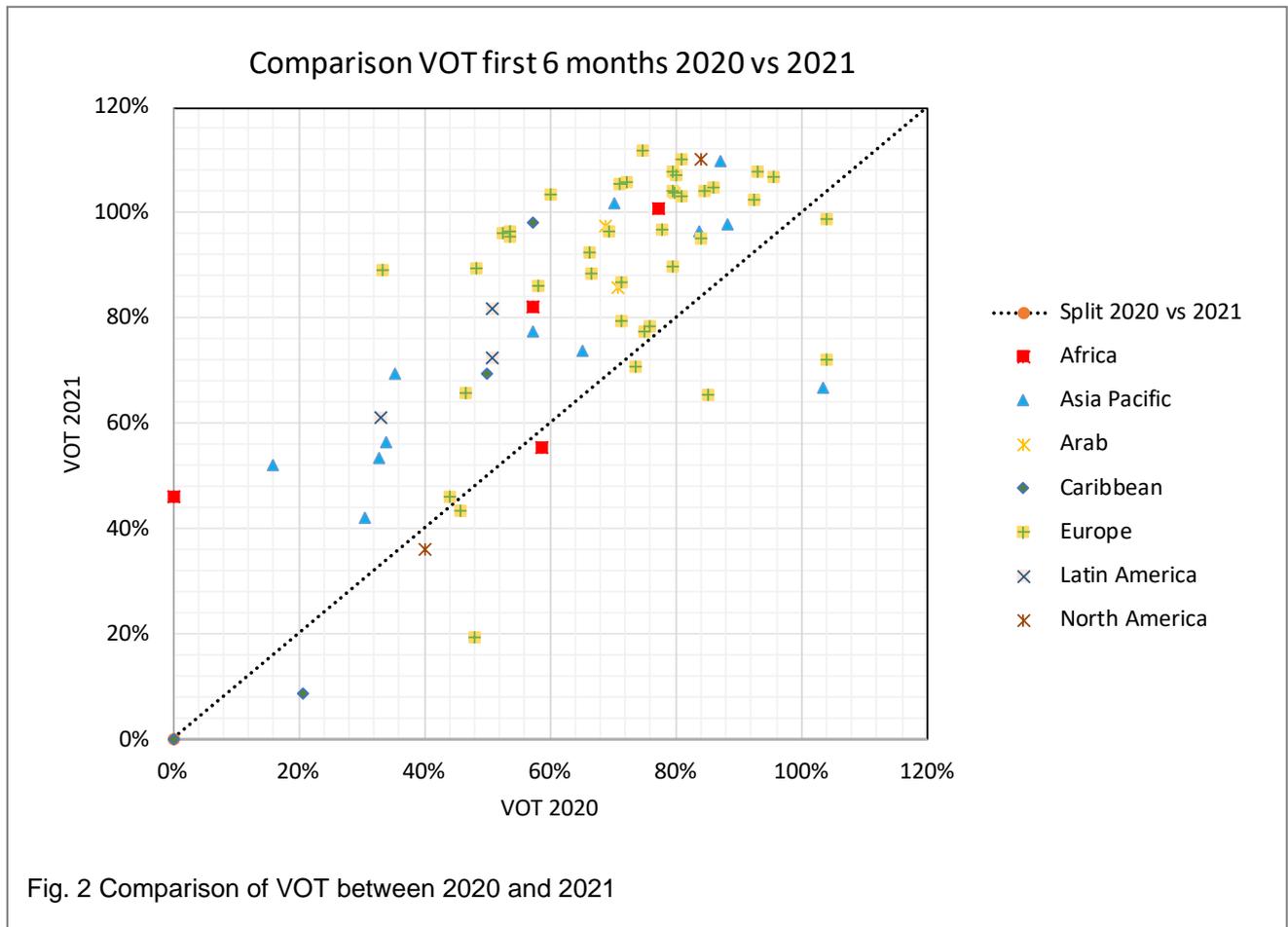


Fig. 2 shows how the average VOT in the six-month window has changed for each measured country between 2020 (value in the horizontal axis) and 2021 (value in the vertical axis). Each country is represented by a dot in a specific colour and shape representing the region (see legend). All countries above the dotted line are countries with an improved VOT in 2021 compared to 2020.

It can be noted that the average VOT has been improving for a large majority of countries. There are still some countries that had their average decreasing in the first six months of 2021 compared to 2020. Six out of the ten countries with a decreased VOT are European.

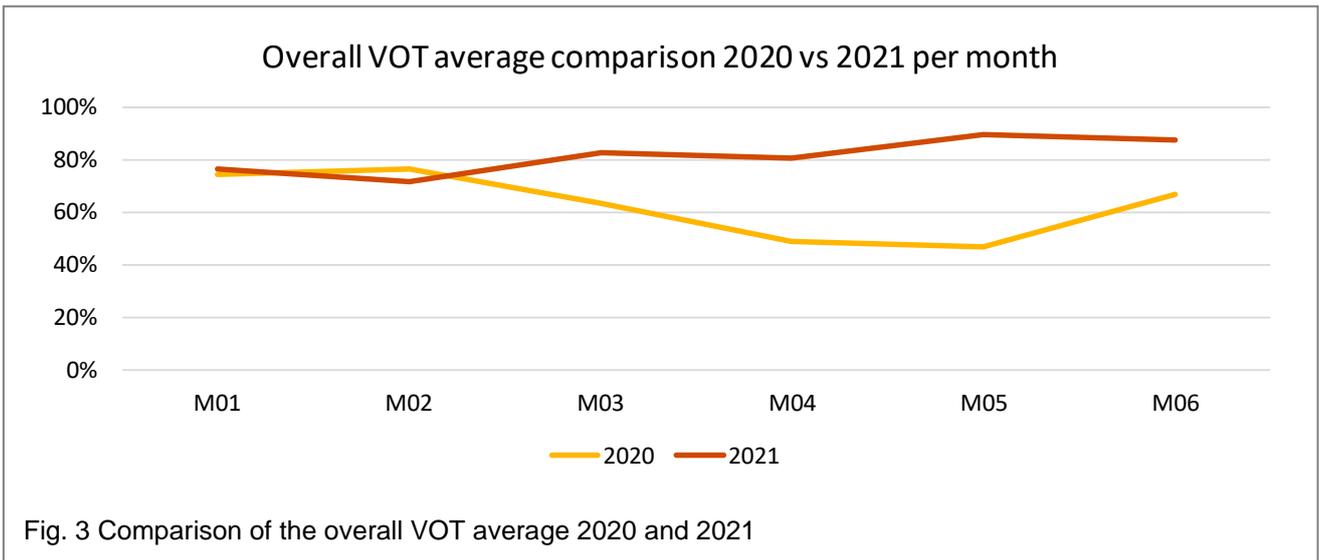


Fig. 3 shows how the overall average VOT has been developing over the six months in scope of the comparison between 2020 and 2021.

The VOT started to decrease after February 2020, when the virus became a pandemic. The decrease continued until May 2020, followed by a rapid recovery.

After February 2021, the recovery of VOT continued almost continuously, remaining at values around 90% during the last two months of the analysed time window.

These figures seem to indicate the reached ability of the measurement to react with a certain level of stability to the different pandemic waves.

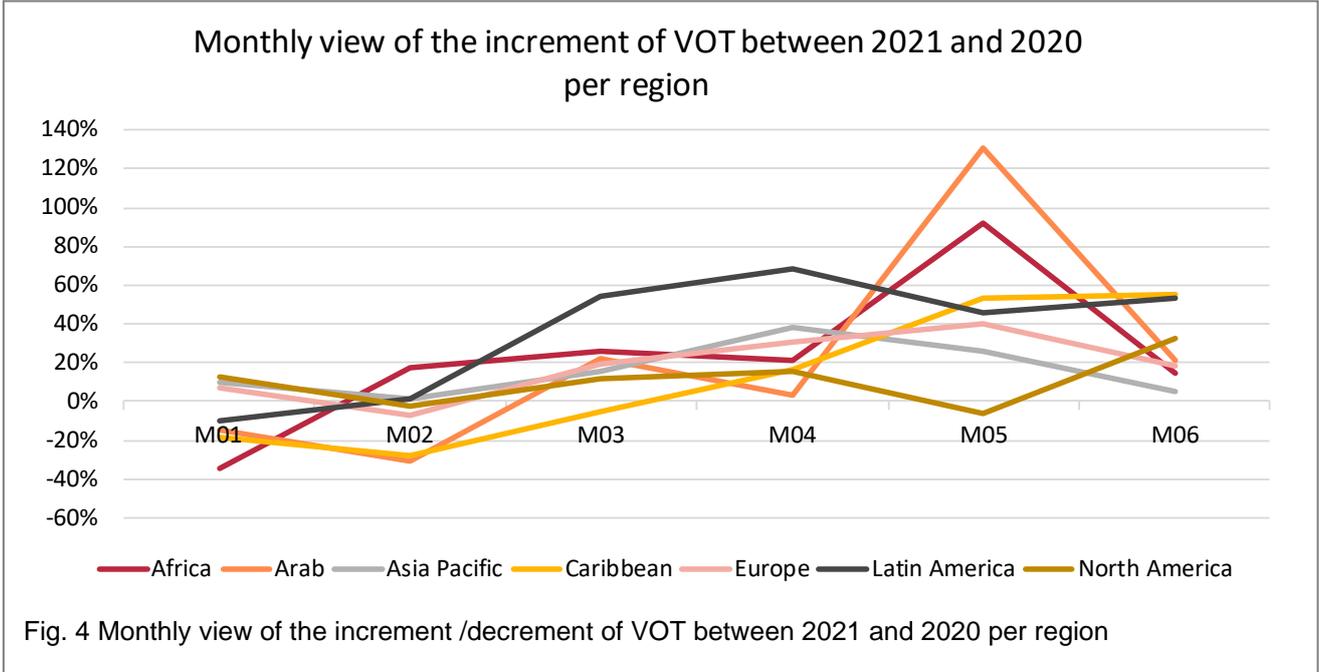


Fig. 4 gives an idea of how the pandemic affected the different regions in waves with different timings. The vertical axis shows the difference of VOT between 2020 and 2021 (VOT in % 2021 – VOT in % 2020).

It is to be noted that changes are mostly positive, meaning that the VOT in the given months was higher in 2021 compared to 2020.

It can also be noted that the waves differ in timing according to regions.

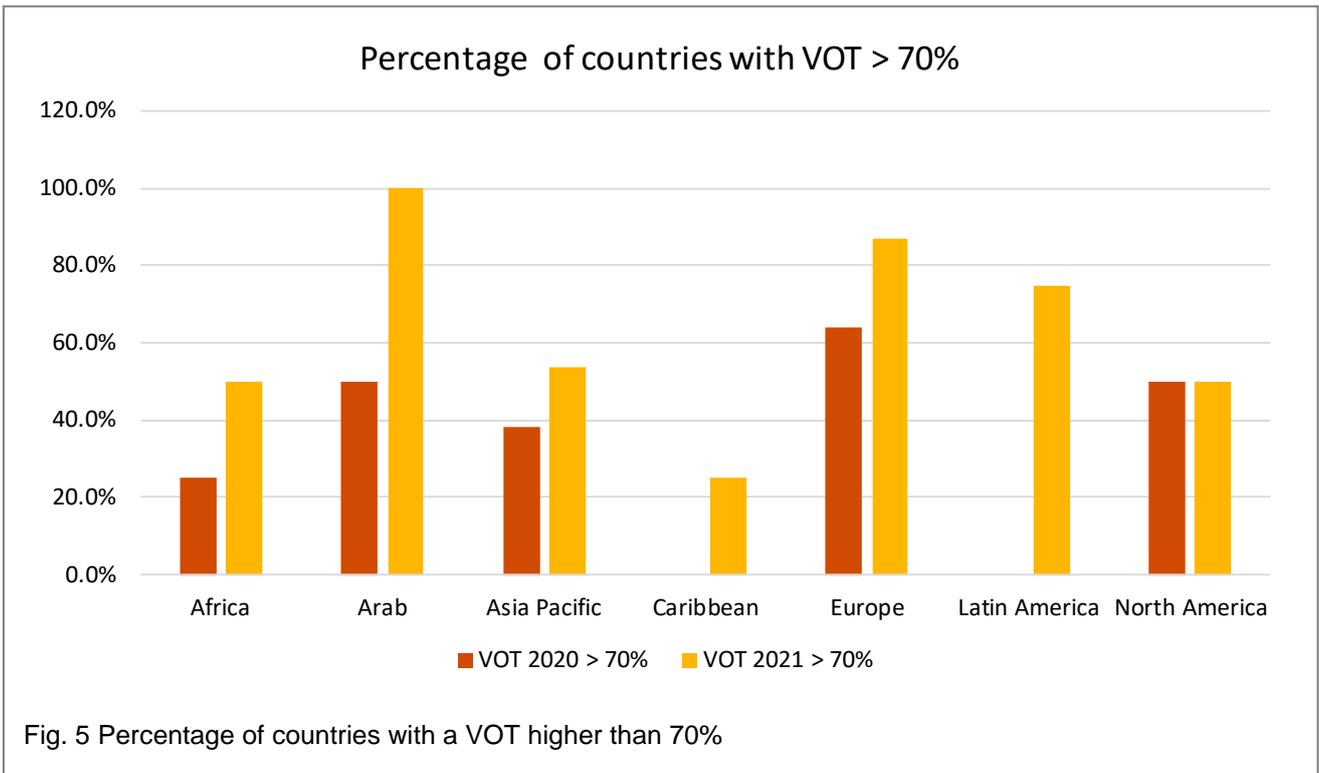


Fig. 5 can be used to understand the VOT at country level. It is to be noted that if the VOT decreases to under 85%, then a re-weighting procedure can be applied. If the VOT decreases to under 70%, then the situation of measurement for this country can be considered critical and could be considered non-compliant.

The figure shows that a large percentage of countries in all regions was below the critical threshold during the first six months of 2020 (35 in total). The number of countries above the threshold has consistently increased during the first month of 2021, reaching 50 countries out of 68 (QS Link countries in 2020).

As stated by the POC, the application of 'deletion for force majeure' can cause the number of valid test items to fall below the acceptable data accuracy threshold needed to use the GMS measurement for the quality link terminal dues. To address this risk, the POC has allowed the use of historical data under certain conditions.

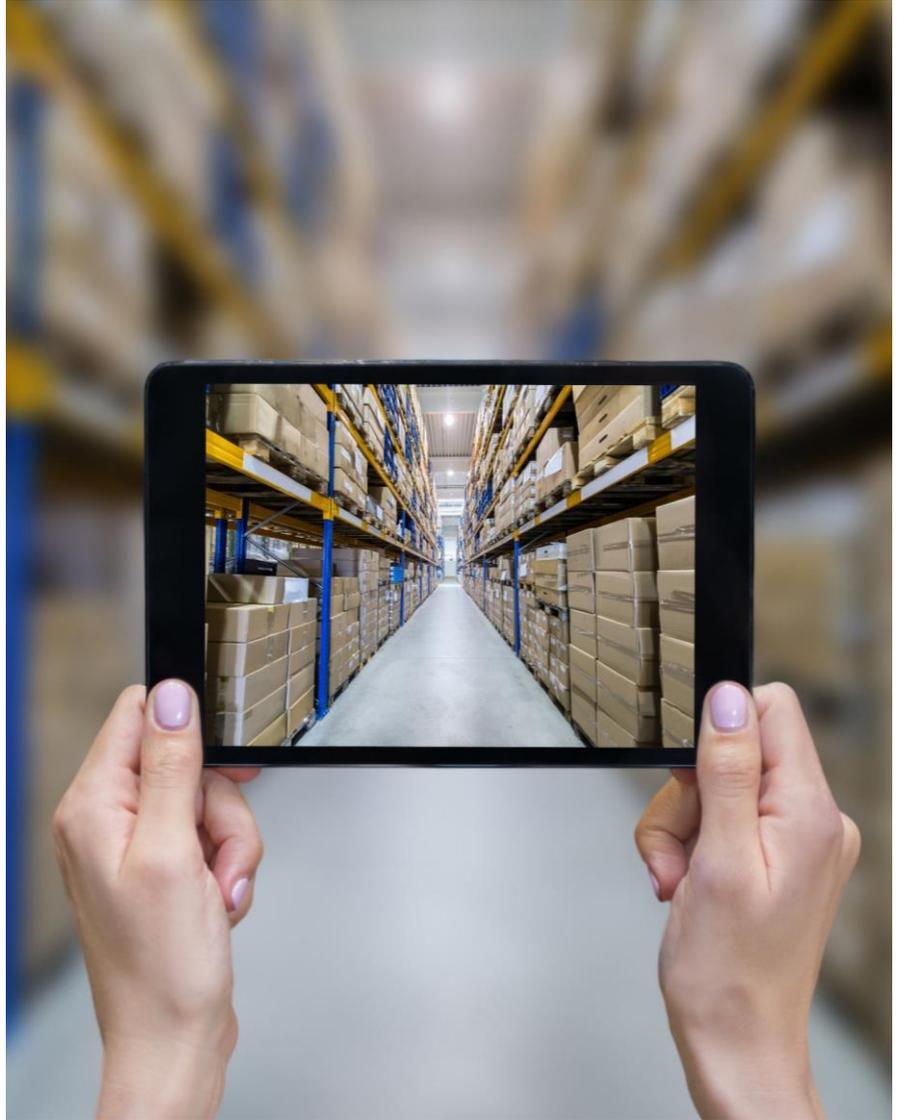
4.4. Points of attention for the POC

The point of attention of 2020 relating to the transition to passive transponders has been considered.

The following point for 2021 could currently represent a compliance issue and we suggest POC to take a decision on how to handle it for the performance measurement.

ID	Title	Description	Suggestion	Status in audit
1	Rules on handling bundling	<p>We noticed that, for the countries in the sample, the number of items that should be considered bundled according to UPU GMS TD (in paragraph 18.1.7 'Bundling on arrival due to operations') is relatively high, in some cases reaching 25% of the items. Since the bundling on the sending side is very limited, the bundling on the inbound side is very likely caused by irregularities in the international transportation.</p> <p>The rule in paragraph 18.1.7 of the UPU GMS TD says that items identified as bundled on arrival due to operations will be excluded from the measurement. Identification can occur on request by the designated operators or on analysis of the MSPs. If automatic checks are introduced by the MSPs there is a potential of a large number of items being duded with severe impact on VOT.</p>	<p>We suggest reconsidering the rule defined in paragraph 18.1.7 of the UPU GMS TD to avoid in future possible discussions in relation to removal of large number of items from the measurement.</p> <p>We also suggest for the performance measurement of 2021 that these items be kept in the calculation to avoid reducing too much the VOT with a larger impact on the measurement compared to the one caused by bundling.</p>	The utilisation of the bundled items for the performance measurement is subject to the decision of the POC.

Annexes



A1 Rating Criteria

Compliance rating criteria

The compliance rating indicated the compliance of the different assessment areas with the UPU GMS TD document.

Non-compliant means a clear violation of the UPU GMS TD document.

Partially compliant means a minor deviation from the UPU GMS TD document with no expected impact on the final measurement results. The significance rating provides indication on the severity and on the priority. Partial compliance can be related to

- a decision to deviate in order to improve quality in certain areas,
- a different interpretation of the UPU GMS TD document or
- a minor mistake in applying the rules.

Compliance rating:

-  Compliant
-  Partially compliant
-  Non-compliant

Significance rating criteria

The significance is an estimation of the impact on the measurement of the identified issue.

- Low means no impact on the measurement results.
- Medium means an impact on the measurement results that should be analysed, but expectation is that the impact does not change the measurement.
- High means that the measurement result is affected, and the implications should be analysed in detail.

Significance rating:

-  Low
-  Medium
-  High