Introduction
The Endocrine Society guideline recommends measurements of plasma free metanephrines for initial biochemical screening of pheochromocytomas and paragangliomas (PPGLs). Liquid chromatography-tandem mass spectrometry (LC-MS/MS) combined with appropriately established reference intervals and pre-analytical precautions is considered the gold standard when measuring plasma metanephrines.

Safe and effective diagnosis and management of patients with PPGLs calls for standardization, or harmonization, of test methods available.

Aim
This international study aimed to establish inter-laboratory comparability of LCMS/MS measurements of plasma free Metanephrine (MN), Normetanephrine (NMN) and 3-Methoxytyramine (3-MT) in external quality assurance (EQA) and patient-derived samples.

The study also compared test interpretations based on the locally employed reference limits used by participating laboratories.

Materials and Methods
Twelve laboratories analysed 125 pooled patient plasma samples which ranged from normal to pathological concentrations. Participation in this study required enrolment in the RCPA-QAP program for plasma metanephrines and results from cycle 20 and 21 (2018) were analysed to further evaluate the agreement of methods.

Agreement in test interpretation was assessed by comparing the number of measurements falling within (non-pathological) and above (pathological) the upper reference limits (URLs) used by each participating laboratory. Agreement between laboratories in the interpretation of non-pathological versus pathological results was assessed using the Fleiss-kappa inter-observer statistical test.

Inter-rater agreement was evaluated for NMN and MN. Results were then normalised to a common reference interval (age specific for NMN) and further evaluated.

Results & Discussion
Regression analysis of test results (Figure 1) revealed satisfactory correlations between laboratories for MN and MN (R ≥ 0.95). The average bias between methods was 1.2% (11.6% to 16.0%) for MN and 0.1% (-18.0% to 9.5%) for NMN. Bland-Altman analysis showed mean differences less than the RCPA-QAP ALP (±15% MN, ±20% MN)
Comparison of 3-MT results between labs showed suboptimal agreement and large proportional biases for at least one third of the labs involved. All labs except labs: A, F, K and M met the RCPA-QAP ALP requirements (±30%).

Passing-Bablok analysis of RCPA-QAP patient samples revealed excellent correlations for all three metabolites. Bias was similar in native patient and RCPA-QAP samples for 3-MT (data not shown) which indicates systematic discordance in the analysis between laboratories (namely imprecision and variability of lower limit of quantification).

The same was not found for NMN and MN which highlights the matrix differences between native and EQA samples – a problem that is often observed with non-commutable QAP materials for many lab tests.

Inter-rater analysis demonstrated good agreement in clinical interpretation for MN, and less so for NMN (Figure 2). Differential adoption of fixed vs age adjusted upper reference limits (URLs) for NMN may contribute to disparities in interpretation.

Following data normalization to a single URL, agreement improved beyond that for normalizing to bias – illustrating the need for not just harmonised analytical methods but also for the adoption of common URL’s.

Conclusion
• Metanephrine shows good agreement across laboratories and large promise for harmonisation of results.
• Normetanephrine is also suitable for harmonisation, however, there is a 16% disagreement in test interpretation – which could be the result of some laboratories using fixed rather than age dependent reference limits.
• 3-MT agreement is suboptimal and requires improvements to analytical methods and reference intervals before harmonisation can occur.

Figure 1: Comparison of analytical test results of NMN, MN and 3-MT in 125 native patient samples measured by LC-MS/MS in 12 laboratories.

Figure 2: Inter-rater agreement in interpretations of test results for NMN (A) and MN (B) using i) the analyzed raw data and participants own cut-offs (open black circles, black line), ii) normalized data and participants own cut-offs (open black triangles, dashed black line) and iii) normalized data in conjunction with normalized cut-offs derived from laboratory A (open grey rectangles, grey line).