



Apple Watch 心率 准确性研究

2026 年 9 月

概述

心率追踪是健康和健身可穿戴设备的重要核心，它不仅为用户带来日常健康和运动方面的洞察，还为多项核心功能提供输入数据，如卡路里消耗估算、最大摄氧量估算（VO₂ max）、经期预测、心率过高或过低时的预警提示等。因此，心率追踪的准确性至关重要，它是实现用户体验中获取有意义洞察的基础。

随着 Apple Watch Series 12 和 Apple Watch Ultra 4 的发布，Apple 引入了健康感测系统。该系统不仅包含了经过升级的电学与光学（光电容积脉搏波描记法，Photoplethysmography，简称 PPG）心率传感器及相关的信号处理算法，还带来了由这项新技术驱动的多项全新功能，例如恢复心率变异性性和“准备指数” app。

本文总结了升级后的光学心率传感系统在可穿戴设备领域中相对于其他设备的准确性。该结论基于一项大型的多中心对标研究，该研究使用心电图（Electrocardiography，ECG）胸带作为参考设备，这是评估可穿戴设备心率准确性的公认标准 [1,2]。本研究包含了超过 1,000 名参与者，在广泛的活动类型中对多种参比设备进行了评估，其充足的样本量足以得出具有统计检验效能的结论。

研究设计

参与者在一只手腕上佩戴 Apple Watch，在另一只手腕上佩戴用于对比的腕戴式设备，同时佩戴 Polar H10 ECG 胸带作为参考设备。手腕的分配采用随机化处理，以确保 Apple Watch 约有 50% 的时间被佩戴在惯用手上。参与研究的腕戴式设备包括：Apple Watch Series 12、Garmin 佳明 Forerunner 970、Google Pixel Watch 4、华为 WATCH 5、三星 Galaxy Watch 8 以及 WHOOP Band 5.0。此外，根据制造商声明的最佳佩戴实践 [3]，参与者还在食指上佩戴了 Oura Ring 5。这些品牌不仅占据了全球智能手表品类约 90% 的市场份额，还包含了目前两款备受瞩目的无屏幕可穿戴设备 [4]。为了确保佩戴合适，研究协调员协助调整每台设备的佩戴，使其既贴合又舒适，并符合制造商的佩戴建议。本研究于 2026 年 7 月和 8 月进行，评估所用设备均为各制造商当时可用的最新型号。

为了确保活动涵盖足够充分的类型以得出具有普适性的结论，参与者完成了不同配速的户外跑步、室内跑步机间歇跑、户外骑行、高强度间歇训练（High-intensity Interval Training, HIIT）以及力量训练，每种运动类型的持续时间在 15 到 20 分钟之间。这一活动组合经过专门挑选，包含了对光学心率性能具有挑战性的条件——例如，与跑步相关的步频锁定（cadence lock）、力量训练和骑行过程中负荷状态下持续抓握及前臂肌肉收缩所造成的信号干扰、以及与 HIIT 相关的快速手腕运动和心率剧烈变化。参与者还完成了一项日常生活测试方案，以评估在静息、案头工作、备餐和步行等活动中的后台心率测量准确性，每项约持续 20 分钟。参与者被允许同时完成运动和日常生活两项测试方案。根据制造商关于避免在举重期间佩戴戒指的建议 [5]，在力量训练中并未对 Oura Ring 进行评估。所有参与者均同意了知情同意书（Informed Consent Form, ICF）中概述的条款。该知情同意书及研究方案获得了 Apple 内部多个审查委员会的批准，委员会成员中包括了来自专门研究伦理、安全性、隐私和数据治理的团队中的代表。

该研究在五个地点进行：两个位于加利福尼亚州库比提诺的 Apple 总部附近，另外三个地点分别位于加利福尼亚州圣地亚哥、得克萨斯州奥斯汀和马来西亚雪兰莪。Apple 负责库比提诺和奥斯汀地点的招募与执行工作，而一家第三方合同研究组织（Contract Research Organization, CRO）负责管理圣地亚哥和雪兰莪的站点。

本研究聚焦于多样化的人群，在性别、年龄、体重指数（Body Mass Index, BMI）和肤色上均有代表性。纳入标准包括：年满 18 岁，能够佩戴腕部、手指和胸带设备，能够承受 30 分钟的体力活动，且符合其他确保参与者安全的标准。排除标准包括：在科技媒体工作、对研究设备存在已知过敏、正在服用 β 受体阻滞剂，或存在任何不宜进行剧烈运动的禁忌症（包括不稳定的心血管疾病）。在设备佩戴部位有纹身或大面积积痣的个体也被排除在外，因为这可能会干扰各制造商设备的读取。

在为每台设备选择要评估的数据流和数据源时，我们优先考虑了三个主要标准。首先，数据流必须是用户在产品体验中可见的，因为本研究的目的是评估用户实际体验到的准确性。其次，数据流必须包含适当的时间戳，以支持与参考胸带进行配对分析。第三，当存在多种采样频率选项时，选择最高频率的数据源，以减少模糊性并降低各制造商引入或省略任何类型平均值处理的影响。

不同设备需要使用不同的方法来获取心率数据。我们为每台设备选择了符合上述所有标准的方法。下表概述了每台设备的方法和心率报告间隔。在多数情况下，本研究采用蓝牙低功耗（Bluetooth Low Energy, BLE）数据流，将厂商心率 app 输出的数据传输至另一部 iPhone 上的数据记录 app。采用该方式是因为其能够同时满足上述全部标准；并非所有厂商都会将高采样精度心率数据写入 HealthKit，且部分厂商亦不支持从其配套 app 中导出此类数据。如果设备支持运动模式，在运动场景测试中开启该模式，在日常生活场景测试中保持关闭。

为获取适用于分析的 Apple Watch 心率数据，除了从 HealthKit 导出的数据外，还使用了一个内部数据记录 app，以采集设备实际向用户呈现的最高频率心率数据。该数据为单次测量值（而非平均值），每 5 秒报告一次；其中，后台心率数据呈现于“表盘复杂功能”（Watch Face Complications）中，运动期间的心率数据则呈现于“体能训练”app 中。出于设备内存管理的考虑，后台心率数据仅每 30 秒写入一次 HealthKit，因此本研究需要使用该内部应用程序，以记录与实际向用户呈现的数据完全一致的心率数据。

尽管各对比设备的心率数据报告时间间隔存在差异，但每个面向用户的数据流均与统一结构的参考设备数据流进行比对评估，以合理评估最终呈现给用户的心率测量体验的准确性。

模式	设备	采集方式	报告间隔
运动模式	Apple Watch Series 12	内部数据记录应用	5 s
	Garmin 佳明 Forerunner 970	BLE	0.25 - 0.5 s
	Google Pixel Watch 4	BLE	2 s
	华为 WATCH 5	BLE	1 s
	Oura Ring 5	HealthKit导出	5 s
	三星 Galaxy Watch 8	通过配套应用导出	1 s
	WHOOP Band 5.0	BLE	1 s
后台模式	Apple Watch Series 12	内部数据记录应用	5 s
	Garmin 佳明 Forerunner 970	BLE	0.5 s
	Google Pixel Watch 4	BLE	2 s
	华为 WATCH 5	BLE	1 s
	Oura Ring 5	HealthKit导出	可变（约每300 s一次）
	三星 Galaxy Watch 8	通过配套应用导出	60 s (步行时为1 s)
参考设备	WHOOP Band 5.0	BLE	1 s
	Polar H10 心电图带	BLE	1 s

表 1. 各设备的心率报告间隔及数据收集方法

本研究预先设定样本量，即针对每种运动类型及每款对比设备，至少纳入 50 名参与者。样本量估算仅适用于运动测试，不包括日常生活测试。样本量计算假设采用双侧检验，显著性水平 $\alpha = 0.05$ ，并针对 6 项设备比较进行 Bonferroni 校正；在此条件下，研究具有 80% 的统计检验效能，可检出平均绝对误差（Mean Absolute Error, MAE，单位：bpm）的标准化配对差异（Cohen’s d_z ）约为 0.5。在假设参与者间 MAE 标准差为 5.0~7.5 bpm、设备间相关系数为 0.5 的情况下，平均 MAE 的最小可检出差异约为 2.5~3.7 bpm，其中户外跑步和骑行约为 2.5 bpm，HIIT 约为 3.7 bpm。上述参数假设仅用于样本量估算，最终统计分析方法如下所述。

方法

本研究对参考设备和测试设备的数据质量检查通过程序化方式执行。对于参考设备，数据需同时满足覆盖度和有效性要求：完整测试时段内的读数覆盖率应达到 $\geq 90\%$ ，单项活动期间的覆盖率应达到 $\geq 70\%$ ；同时，超出 30~220 bpm 范围的测量值比例不得超过 5%。对于测试设备，则要求存在心率数据流且没有极端值（即超出 20~240 bpm 范围的测量值；由于制造商内部报告设定的心率报告范围未公开，本研究采用了相对宽泛的 20-240 bpm 作为极端值判定范围）。若某一单项活动未通过上述任一项质量检查，则该项活动的数据将从受影响的相应设备比较中排除。另外，参考设备中超出 30~240 bpm 范围的单个采样点亦从分析中排除。

各次数据采集的起始时间根据可用时间戳信息与参考胸带的起始时间进行匹配。为了确保设备使用网络时间协议（Network Time Protocol, NTP）进行准确的时间同步，每台设备（包括 Apple Watch 和 Polar H10 胸带）均通过蓝牙®与各自专属的、已连接 Wi-Fi 的配套手机完成配对。研究协调员在数据收集开始之前确认所有设备对齐程序均已完成，并在数据导出之时再次确认完成对齐后，将各设备的每个测量值与参考胸带中时间戳最近的采样点进行最近邻匹配（nearest-neighbor matching），从而能够在单采样点层面对各厂商面向用户设计的心率测量体验相对于参考标准的准确性进行评估。

三星 Galaxy Watch 8 在日常活动（后台心率）场景下的数据需要采取不同的处理方式，因为导出数据并未提供对应的时间戳，仅提供了开始和结束时间戳、最低和最高心率，以及一个未记录推导方式的心率值。将该值与时间段开始时、时间段结束时的参考值以及时间段内参考值的平均值进行对比测试，结果显示与时间段平均值的一致性最高。因此，在这项日常活动场景下记录的心率，是与胸带参考在每个时间段内的平均值进行比较的。

每次测试的准确性均采用均方根误差（Root Mean Square Error, RMSE，单位：bpm）和平均绝对误差（Mean Absolute Error, MAE，单位：bpm）进行评估。二者均为比较性准确度评价中的常用指标，其中 RMSE 对识别偶发但显著的误差更为敏感。Apple Watch 在每个时段的每个指标（RMSE 和 MAE）都与对比设备的指标进行了直接比较。我们还将各设备在各时段的心率均值与参考胸带进行了比较，然后对标评估了各设备的心率均值误差。这是一个时段总结的指标，这对于没有屏幕的设备尤为重要——这类设备在不连接智能手机的情况下无法向用户提供实时心率数据。

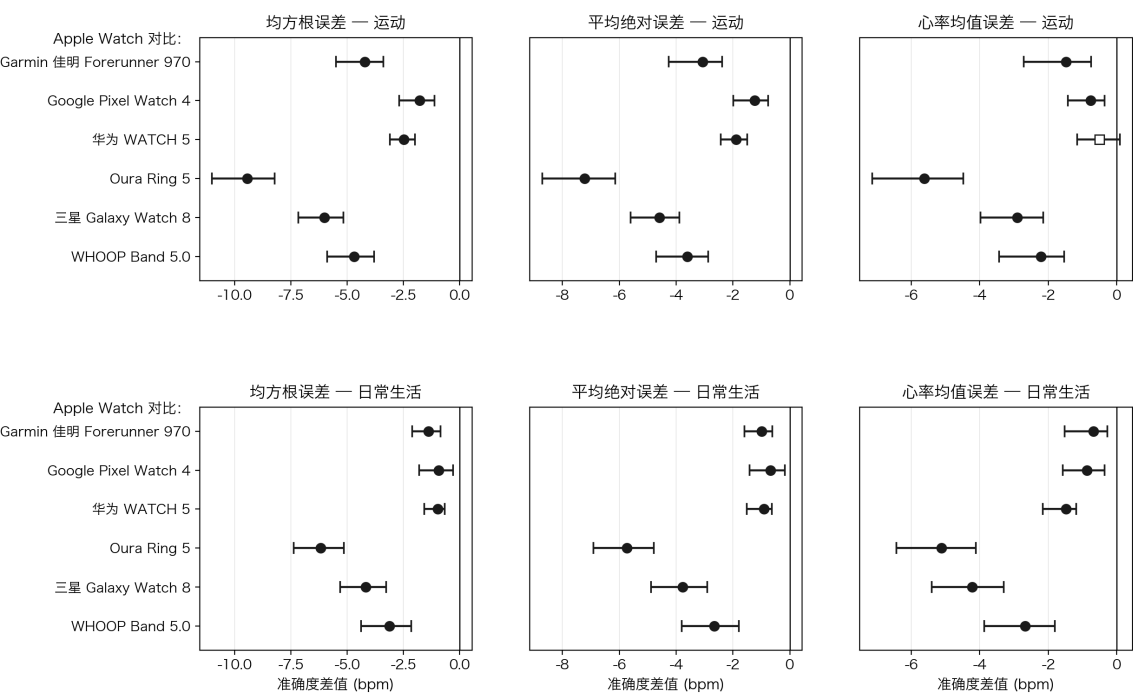
本研究采用线性混合模型（Linear Mixed Model, LMM）来对每位参与者和每项活动的每个准确性指标的配对差异（Apple Watch - 对比设备）进行建模。模型将运动类型作为固定效应，并引入受试者随机截距以校正重复测量导致的受试者内相关。

所报告的总体准确性差异定义为各活动类型平均差异的等权平均值。考虑到多款设备比较所带来的多重性问题，本研究同时计算按受试者聚类的 99% 偏差校正与加速（bias-corrected and accelerated, BCa）自助法置信区间。对于平均误差，首先计算每次测试的绝对误差，再对这些绝对误差取平均值。因此，所报告的设备间差异实质上为平均绝对误差之差。所有报告指标均表示平均误差差值，单位为 bpm。由于差值定义为“Apple Watch 误差减去对比设备误差”，因此负值表示 Apple Watch 的测量误差低于相应对比设备。

结果

本研究共招募了 1460 名参与者，在肤色、性别、BMI 和年龄方面均有代表性。其中，1254 人在最终分析中贡献了至少一组 Apple Watch 与对比设备的配对测量数据。在这些人中，59% 是在第三方 CRO 管理的站点招募的，20% 的人的 Fitzpatrick 肤色分级为 V 或 VI，35% 为女性，平均年龄为 39 ± 12 岁，平均 BMI 为 $26 \pm 5 \text{ kg/m}^2$ 。

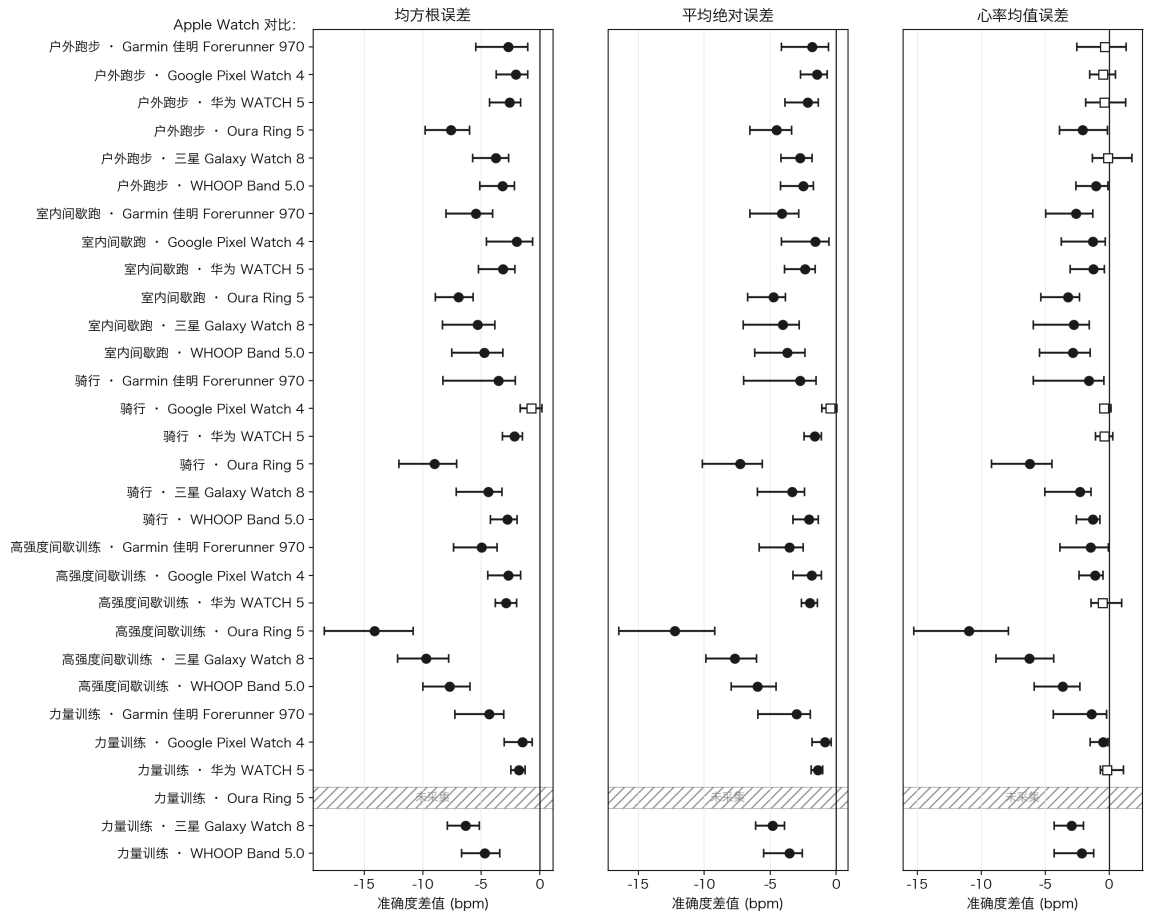
对于 RMSE 和 MAE 而言，相比所有其他设备，Apple Watch 在整体运动和整体日常生活（后台心率）测试方案中均表现出了具有统计学意义的更高准确度。对心率均值而言，在 12 次比较中，有 11 次 Apple Watch 也展现出更高的准确性；唯一的例外是与华为 WATCH 5 在整体运动中的比较，Apple Watch 表现出了更好的数值准确性，但在差异未达到统计学显著性。结果详情见图 1 至图 3，涵盖了整体结果、单项运动结果以及单项日常活动结果。



误差差值及其 99% 置信区间。实心圆表示设备间存在统计学显著差异；空心方块表示差异在统计学上未达到确定性判定水准（无显著统计学差异）。

图 1：整体结果（Apple Watch 测量误差减去对比设备测量误差）

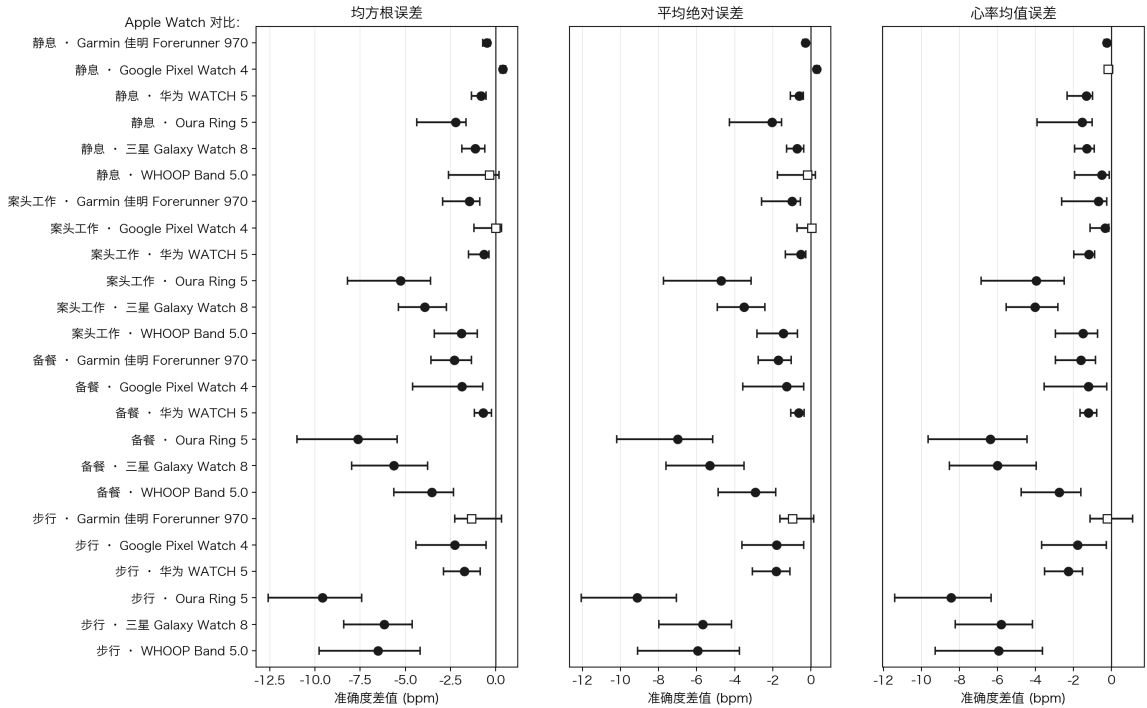
在除一项外的所有单项运动（骑行、户外跑步、室内间歇跑、HIIT、力量训练）中，Apple Watch 针对所有设备的 RMSE 和 MAE 指标均展现出了统计学上更好的准确性；唯一的例外是骑行时与 Pixel Watch 4 的比较，该差异在统计学上未达到确定性判定水准（无显著统计学差异）。在单次测试心率均值方面，Apple Watch 在大多数活动中相较于大多数设备展现了更高的准确性，但在 8 项对比中，差异在统计学上未达到确定性判定水准。



误差差值及其 99% 置信区间。实心圆表示设备间存在统计学显著差异；空心方块表示差异在统计学上未达到确定性判定水准（无显著统计学差异）。

图 2：各项运动结果（Apple Watch 测量误差减去对比设备测量误差）

对于日常生活测试方案中的各项活动，Apple Watch 在几乎所有活动中、针对几乎所有设备的所有指标均表现出了更好的准确性，仅有少数例外。在静息、案头工作和步行状态下，有 8 项比较在统计学上未达到确定性判定水准。仅在静息条件下，Pixel Watch 4 的 RMSE 和 MAE 准确度更高，但优势幅度不足 0.5 bpm。



误差差值及其 99% 置信区间。实心圆表示设备间存在统计学显著差异；空心方块表示差异在统计学上未达到确定性判定水准（无显著统计学差异）。

图 3：各项日常生活结果（Apple Watch 测量误差减去对比设备测量误差）

在涵盖运动和日常生活测试方案的所有单独活动中，共进行了 159 次特定活动与衡量指标的对比。Apple Watch 在其中的 139 次比较中在 99% 置信水平下展现了更高的准确性，18 次对比未能在统计学上达到确定性判定水准，另有 2 项比较中，对比设备的准确性更高。若不考虑统计学显著性、仅评估点估计值，比较结果在 159 次中有 155 次支持 Apple Watch。

结论

在本项具有统计检验效能的大规模研究中，Apple Watch 在心率测量准确性方面优于所有纳入比较的设备。研究广泛覆盖了智能手表、智能戒指及健身手环用户常见的多种运动和日常活动场景。在本研究所评估的可穿戴设备中，Apple Watch 为用户提供了最高的心率测量准确性。

参考文献

1. Chung V, Chopin L, Karadayi J, Grèzes J. 2026. "Validity of the Polar H10 for Continuous Measures of Heart Rate and Heart Rate Synchrony Analysis." *Sensors (Basel)*. Jan; 26(3):855. [pmc.ncbi.nlm.nih.gov/articles/PMC12899868/](https://pubmed.ncbi.nlm.nih.gov/articles/PMC12899868/).
2. Schaffarczyk M, Rogers B, Reer R, Gronwald T. 2022. "Validity of the Polar H10 Sensor for Heart Rate Variability Analysis during Resting State and Incremental Exercise in Recreational Men and Women." *Sensors (Basel)*. Aug; 22(17):6536. [pmc.ncbi.nlm.nih.gov/articles/PMC9459793/](https://pubmed.ncbi.nlm.nih.gov/articles/PMC9459793/).
3. Oura. "How to Choose the Right Oura Ring Size." Oura Support. Accessed August 2026. support.ouraring.com/hc/en-us/articles/360025590653-How-to-Choose-the-Right-Oura-Ring-Size.
4. Omdia. "Wearable Band Sell-In Shipments, 2Q26."
5. Oura. "General FAQs." Oura Support. Accessed August 2026. support.ouraring.com/hc/en-us/articles/4408961184147-General-FAQs.