Introduction

This Annex explains the data available from the electrocardiogram (ECG) export. The reader should consult the main data-dictionary[1] for a description of the screening protocol itself, explanations of formatting conventions, sentinel values, and other technical information.

The Airwave team are not experts in this subject area, and so we are interested in feedback from specialists that may help us improve or clarify areas of uncertainty.

Title	Data Dictionary, Annex C
Subject	Electrocardiogram (ECG) Export
Version	3.0
Author	Heard, Andy H, Database Manager at Imperial College London. a.heard@imperial.ac.uk
Published	31/10/2024

Changes since Previous Version

1. An additional set of data – "GRIASC" files – are now available. These allow researchers to analyse the raw trace data in detail. We additionally provide an index of these files.
2. Some minor field changes have been made
3. Three data linkage errors from previous revisions have been corrected.

ECG Protocol

We carried out ECGs on most participants at baseline, and some during follow-up. The follow-up programme completed in 2023 but no ECGs were conducted after March 2020. Initially this was because of practical difficulties in taking an ECG during the pandemic; and when we resumed post-pandemic, a funding shortfall required us to cut this protocol.

ECGs were collected on machines recommended by Professor Peter McFarlane's CARE (Computer Assisted Reporting of Electrocardiograms) team at Glasgow University. We uploaded data to CARE daily and they returned an interpretation to us, usually within 6-weeks. We are grateful to Professor McFarlane and his team, Louise Inglis and Shahid Latif, for their long-standing contribution to the Study.

Interpretation Systems

The interpretation systems used by CARE evolved over time. The first version was used from the 2004 until June 2011; a later version from June 2009 until May 2015 (note the overlap, which resulted from compatibility issues with older ECG carts), and the most recent version from November 2015 onwards. Later systems yielded more analytics.

A small number of results were interpreted manually from the paper trace. This happened when data being transmitted to CARE was lost because of technical or procedural faults. Very limited interpretation of these ECGs is available.

Interpretation of the ECG was returned to us in two parts:

- A summary statement aimed at advising the participant and / or their GP whether there was anything that merited further clinical investigation.

[1] https://www.police-health.org.uk/researchers

- A detailed dataset intended for use by researchers.

The summary information was one of NORMAL, BORDERLINE NORMAL, BORDERLINE ABNORMAL, and ABNORMAL. Optionally, there was coding for bradycardia or tachycardia. Participants were generally advised to consult their GPs if they were either of the abnormal diagnoses. In these cases, participants and / or their GPs were also sent a copy of the ECG trace.

ECG Detailed Results

This dataset is made up from the following files:

- **Geometry**: a table of records for each ECG with its summary results.
- **Group Codes**: structured commentary on the ECG using codes proprietary to Glasgow CARE.
- **Minnesota Codes**. We understand these to be a structured commentary similar to Group Codes, but based on an internationally standardised set of codes.
- **GRIASC Files,** which we understand to be a dataset that codes each ECG's signal in sufficient detail to reproduce the waveform. Each file represents a single trace and was provided to use by CARE. For more details, please refer to the GRIASC File Specification[2].

Group Codes

Each Group Code is represented as three values: a group number (2-digits), a statement type (2-digits) and statement code (5-digits). Taken together, they form a compound key within a table of narrative interpretations. We present here the narrative interpretation only, though raw group-codes are available if required (contact the Airwave team for details).

Statements are intended to be read in order, and the succeeding statement can sometimes refer to the preceding statement. For example, PROBABLE NORMAL VARIANT may appear more than once, and in each case it refers to the previous statement.

Minnesota Codes

Each Minnesota Code consists of a 3-digit number that corresponds to a narrative interpretation. As we have no easy access to a lookup table, we have exported the codes we received from CARE in the format given to us. The hyphens were removed by CARE, though we understand the normal convention is to keep them in.

GRIASC Files

The GRIASC-format files code the raw signal for each ECG. The files presented here were extracted from CARE's systems and we have made no modifications to them. For more details, please refer to the GRIASC File Specification.

Linkage

The various datasets received from CARE were linked to our own clinic records by the Study team. Here is some background material for anyone curious about this process.

Firstly, you may recall the structure of the cohort's identifiers from the main Data Dictionary. Each participant has a unique identifier - part_id – that can be used to link throughout the cohort[3]. When

[2] GRIASC-file-specification-v1.4.pdf

[3] Because we have used part_id in communications with participants, we are gradually replacing it by an equivalent, subject_id. Both identifiers are provided in this version of the Geometry table.

a participant attended a clinic, they were assigned a visit-id that links together all the data for that clinic attendance. We called this visit-id the "barcode", a convention that was less than ideal, in retrospect. However, it links the ECG to laboratory assays, metabolomics results and surveys. Each participant has zero, one or more barcodes, all from different visits.

Each visit will typically result in just one ECG, identified by its own ecg_id. Because a trace was sometimes taken more than once, there can be more than one ecg_id per visit.

Linking GRIASC files with the tabular ECG tables is via the barcode reported plus the recording time. Errors in data entry were made at the clinic, and these been corrected. ***You should linking using barcode_true, not the barcode found in the GRIASC file***. Please get in touch with us if you think our linkage is faulty.

There is a separate nuance for the post June-2015 ECGs, which are all follow-up. Because CARE wanted to link these new ECGs with the baseline ECG they already had on file, we agreed that the follow-up ECG would identified using the barcode collected at their original visit. We then sorted the linkage at our end. For this reason, barcode_reported and barcode_true differ for all follow-up ECGs.

Figure 1 attempts to show the relationship for a general case.

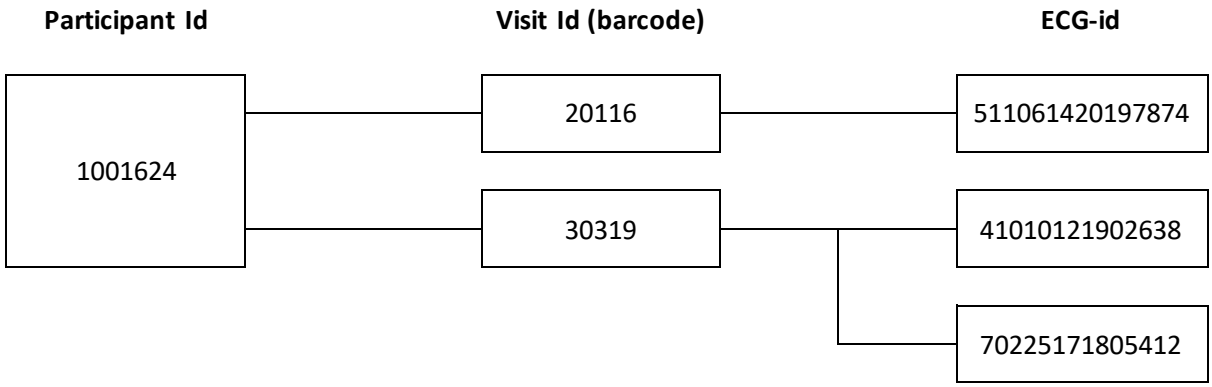

Figure 1

Sentinel Values

In other datasets, we use a sentinel value to explain why certain values are missing (ex-protocol, not-found, not collected etc.). Negative numbers are used as sentinels; but here, negatives are valid, so we are leaving nulls in this extract. The most likely reason for there being no ECG is equipment or transmission failure. In some early cases, we have summary results but no detailed information.

Tables of Variables

The tables below describe each of the variables in the extracted files.

Trace Geometry

This file provides top level summary information on the ECG.

Label	Type	Commentary
ecg_id	NUMBER (17)	This is a unique ID generated by the ECG management system. It appears to be generated from the recording date, time and device name.

Label	Type	Commentary
barcode	NUMBER (5)	Health-screening visit identifier.
seq#	INTEGER(1)	Which visit this is: first = 1, second = 2, etc.
part_id	NUMBER (7)	Pseudonymous identifier that is unique for each participant.
subject_id	INTEGER	Anonymised identifier for the participant that will replace part_id in the next version of this export.
when_recorded	DATETIME	A timestamp from the ECG machine.
interpretation_system	STRING(1)	System version of the interpretation system used. See explanation above.
summary_interpretation	STRING	The summary interpretation. We were advised by Professor McFarlane that only summary reports including the word "ABNORMAL" suggest clinical investigation.
lvh_interpretation	STRING	A statement relating to any ventricular hypertrophy diagnosed. This is only available when interpretation = 3; otherwise "EX-PROTOCOL".
minnesota_codes_ct	NUMBER	Quantity of Minnesota Codes reported
group_codes_ct	NUMBER	Quantity of Group Codes reported
griasc_filename	STRING	Name of the GRIASC file coding the raw signal.
linkage_remarks	STRING	Any comments on special processing to link this ECG.

Table 1: Geometry Fields

The following numeric fields describe the ECG geometry.

- heart_rate
- p_axis;
- qrs_axis;
- t_axis;
- qrs_duration;
- pr_interval;
- qt_interval;
- qtc_interval.

These numeric results are available when interpretation_system = 2 or 3. There is some background material on these analytics in Appendix A.

- p_duration
- lvmi_rautaharju
- lvmi_f_huwez
- cornell_index
- cornell_product
- sokolow_lyon
- qrs_voltage_sum
- qrs_voltage_prod

Group Codes

The following table lists the Group Codes output for each ECG.

Label	Type	Commentary
ecg_id	STRING (25)	The unique ID generated by the ECG management system.
barcode	NUMBER (5)	Health-screening visit identifier.
part_id	NUMBER (7)	Pseudonymous identifier that is unique for each participant.
subject_id	INTEGER	Anonymised identifier for the participant that will replace part_id in the next version of this export.
statement_number	NUMBER(2)	Statement number within group.
text	STRING	Textual interpretation.

Table 2: Group Codes

Minnesota Codes

The following table lists the Minnesota codes output for each ECG.

Label	Type	Commentary
ecg_id	STRING (25)	The unique ID generated by the ECG management system.
barcode	NUMBER (5)	Health-screening visit identifier.
part_id	NUMBER (7)	Pseudonymous identifier that is unique for each participant.
subject_id	INTEGER	Anonymised identifier for the participant that will replace part_id in the next version of this export.
group_ident	STRING(2)	Group name / number.
statement_number	NUMBER(2)	Statement number within group.
non_site_specific	NUMBER(3)	Minnesota code for non-site specific groups.
anterolateral	NUMBER(3)	Minnesota code for anterolateral sites.
posterior	NUMBER(3)	Minnesota code for posterior sites.
anterior	NUMBER(3)	Minnesota code for anterior sites.

Table 3: Minnesota Codes

GRIASC Files Index

The following table provides summary details of each GRIASC file.

Label	Type	Commentary
filename	STRING (15)	Filename
barcode	NUMBER (5)	Health-screening visit identifier.

Label	Type	Commentary
when_recorded	DATETIME	The ECG's timestamp according to the file. These should match the same variable in the Geometry file, but for reasons unknown there are sometimes differences.
barcode_reported	NUMBER (5)	Health-screening visit identifier as reported in the file.
barcode_true	NUMBER (5)	Actual health-screening visit identifier following linkage translation and corrections.
linkage_remarks	STRING	Any comments on special processing to link this ECG.

Table 4: GRIAS Files Index

Variables Absent at this Release

Two variables not exported are:

- ecg_cart: available for 16% of results, this is an identifier for the machine used.
- ecg_comment: available for 3% of results, this is free format text relating to the procedure (e.g. "lead 1 won't stick to patient"), or participants' comments. Because of the risk of disclosing personal data, we have not exported these.

If you want access to these data, please get in touch with us.

We also received for most ECGs a copy of the trace in PDF form. These documents seem to be aimed at a clinical audience, and include participant identifiers. We understand that the GRIASC files allow researchers to recreate the trace if required. If you still believe you could benefit from access to the PDF traces, please contact us.

Files' External Characteristics

This is revision 3 of this extract, extracted in October 2024. The following table relates to the extract available on Imperial College servers.

File	Files	Records	Size (Bytes)	CRC32
Geometry	1	61,736	10,859,305	50848E0F
Group Codes	1	270,348	18,460,890	C74F7BCD
Minnesota Codes	1	119,320	6,923,295	4B15DECC
GRIASC Index	1	61,592	3,390,032	678B7CA6
GRIASC Files	61,643		11,464,240,871	AA2492B6-00007840

Table 5: Files' External Characteristics

Appendix A (LVMI and Voltage Based Indices)

The description below was provided to us by Glasgow CARE in August 2011.

LVMI Rautaharju

Described in [Rautaharju PM, et al; Am. J. Card. 1988; 62:59-66]. The units are g/m^2.

Males: -36.4 + 0.01 RV5 + 0.02 SV1 + 0.028SIII + 0.182 (T-) V6 - 0.148 (T+) aVR + 1.049 QRSd

Females: 88.5 + 0.018RV5 + 0.053 SV5 – 0.112 SI + 0.108 (T+) V1 + 1.7 (T-) aVF – 0.094 (T+) V6

Where Q, R, S voltages are all positive (μV) and S=max (Q,S) and R=max(R,R').

The equation for black females is different from the one above, and has not been used in this extract.

LVMI F. Huwez

Described in [Huwez F; PhD thesis, University of Glasgow, 1990]. The units are g/m^2.

Males: 99.6615 + 0.03358 SV5 + 0.10121 (T+) V1 – 0.4231 (T-)V1 + 0.12692 (T-)V6.

Females: 115.1371 – 0.04284 RIII – 0.02134 SV3 + 0.4927 STV1 + 3.1725.

Cornell Index

Described in [Casale PN, et al; J. Am. Coll. Cardiol.; 1985; 6: 572-580] and [Casale PN, et al; Circulation; 1987; 75: 565-572]. The units are μV.

The Cornell index is the sum of the maximum R (or R') amplitude in aVL and the maximum of absolute S or absolute Q amplitude in V3. For females, this value is increased by 600μV.

Males: RaVL + |SV3|

Females: RaVL + |SV3|+ 600

Cornell Product

Described in [Molloy TJ, et al; J. Am. Coll. Cardiol.; 1992; 20: 1180-1186-580] and [Okin PM, et al; J. Am. Coll. Cardiol.; 1995; 25: 417-423]. The units are μV.S.

This Cornell Product is defined as Cornell Index x QRS duration.

Sokolow-Lyon Index

Described in [Sokolow M, Lyon TP et al; Am. Heart J. 1949; 37: 161-186]. The units are μV.

The Sokolow-Lyon index is the sum of the maximum of absolute S or absolute Q amplitude in V1 and the maximum R (or R') amplitude in V5 or V6: |SV1| + maxRV5/6.

QRS Voltage Sum

Described in [Okin PM, et al; Hypertension 1998; 31: 937-942]. The units are μV.

QRS Voltage Sum is the sum of absolute values of QRS amplitudes in each lead.

QRS Voltage Product

Described in [Okin PM, et al; Hypertension 1998; 31: 937-942]. The units are μV.S.

The QRS Voltage Product is defined is the sum of the products of QRS duration and the absolute values of QRS amplitudes in each lead.