

Data collection field guide

A guidebook for collecting data from routine clinical records

Who is this guide for?

This guide is for data collectors, the people responsible for extracting data from paper clinical records. It summarizes the main steps and requirements to extract the baseline data from clinical records. If you will be extracting data, you should read through this guide.

What do you need?

1. A method to record extracted data. This could be:
 - a. A laptop without Internet connection (for use with the offline data collection form).
 - b. Paper copies of the data extraction form. You will need one form per record.
2. A mobile smartphone or computer with Internet access to randomly generate numbers (for example, the [Google random number generator](#)).
 - a. If you do not have access to this, you will need to determine an appropriate method to randomly sample patient records. This could be done using a spreadsheet or through a previously made random number list, for example.
3. A partner with whom to do the extraction.
4. A coin to flip to help make random selections.

Remember to bring

- A mobile smartphone with data connection (if possible).
- A phone charger or laptop charger (so you do not run out of battery).
- Pens to write with.
- A data extraction method (online form, offline form or paper extraction form).

What to do when you arrive

1. Randomly select a doctor at the clinic

If the clinic has only one doctor, you can skip this step. Further, if the study coordinators are requesting that you extract data from patients based on health facilities (rather than doctors), you can skip this step.

If the clinic has more than one doctor, you need to randomly select a doctor whose files you will use.

To do this you will need a coin. If there are two doctors, randomly assign one doctor to each side of the coin. Flip the coin in the air and choose whichever doctor corresponds to the side of the coin that landed upward. If there are three or more doctors, pick any two doctors, flip the coin, pick a winner, and then repeat using the winning doctor and the third doctor until you have picked one doctor.

If you have not reached your required sample size after extracting information from the doctor you have randomly selected in step 1, randomly select a doctor from the same clinic. Repeat this process until you have reached your target sample size or have extracted data from all of the doctors' patients that meet the inclusion criteria.

Worked example

- You have been tasked with travelling to a health facility (Clinic A) to extract data from 50 adult patients (18 years or older) who have visited the clinic at least once in the previous year.
- Clinic A has seven doctors, each with their own patient lists.
- You will randomly select, using the coin method described above, which doctor to include in the sample.
- You select by random chance doctor number 4.
- You find this doctor's patient register and randomly select patient records in accordance with the next step.

2. Select the patient records

You use the patient register of the doctor you selected in step 1. Depending on the setting, there may be a central list of patients or multiple lists based on patient demography (such as age or diagnosis). You confirm with the study coordinators the register or list from which you are to randomly sample patients.

Count the total number of patients on the register. If the number of patients is less than your sample size, you can include all the patients. If the number of patients is greater than your sample size, proceed to randomly select patients with the method described below.

Use the [Google random number generator](#) or equivalent product or software. Enter the total number (n) of patients on the list as the MAX value. Click GENERATE. This will give you a random number between 1 and n . Use this number to count down the list of patients on the register, find the chart of the chosen number. If you cannot access the Google random number generator, speak with the project coordinators about alternative methods.

Once you find the file, check to make sure the patient meets the inclusion criteria. If they meet the inclusion criteria, continue to the next step; if they do not, put the file back and select the next file from the random chart selection list.

Worked example: illustration of the method used to randomly sample patient records from a hypertension register in Tajikistan

Example Hypertension Register			Dr. B.P. Ressure.
Name	Sex	Date of Diagnosis	
1. J. Doe	Male	DD/MM/YYYY	
2. J. Doe	Female	DD/MM/YYYY	
3. J. Doe	Male	DD/MM/YYYY	
4. J. Doe	Male	DD/MM/YYYY	
5. J. Doe	Female	DD/MM/YYYY	
6. J. Doe	Female	DD/MM/YYYY	
7. J. Doe	Female	DD/MM/YYYY	
8. J. Doe	Male	DD/MM/YYYY	

Step One

- Count number of patients documented on the hypertension register, by gender

Step Two

- Use a computerized random number generator to create a random list of records to sample, separately for men and women

Step Three

- Locate the chart of a randomly selected patient and check for inclusion criteria
 - ✓ Was at least age 18 or older 12 months ago
 - ✓ Has visited the clinic at least once in the last 12 months

Step Four

- Repeat until sample size is reached for both gender samples

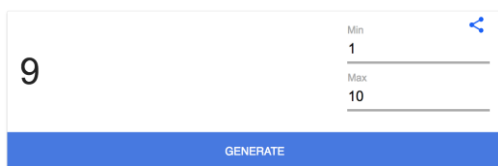
Source: Collins DRJ, Laatikainen T, Shoismatuloeva M, Mahmudzoha I, Rahimov Z, Sultonova D et al. Evaluation and pilot implementation of essential interventions for the management of hypertension and prevention of cardiovascular diseases in primary health care in the Republic of Tajikistan [version 1; peer review: 1 approved]. F1000Research. 2019;8:1639 (<https://doi.org/10.12688/f1000research.20234.1>, accessed 25 May 2020).

To ensure the minimum age to meet the inclusion criteria, you may need to determine what the patient's age was in the preceding year. For example, if the patient is required to be 18 years or older, you must look through the file and determine whether the patient has visited the clinic within the past 12 months. Use your own judgement to determine this and look at the dates for the most recent clinical information. If they have visited the clinic within the past 12 months, keep the record and include it for data extraction. If they have not, put the file back and select the next file from the random chart selection list.

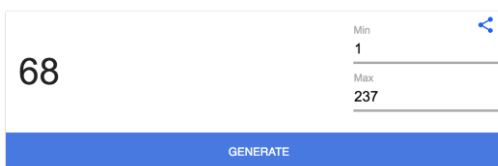
Worked example

→ After finding the hypertension register, you count the total number of patients to be 237.

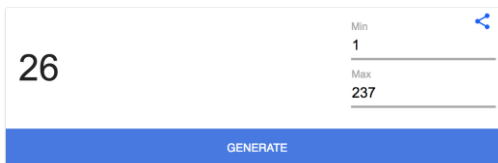
- ◆ Open the [Google random number generator](#) on your mobile smart phone. The screen will look like this.



→ Change the MAX value to 237, so that it looks like this:



→ Click generate, and the application will randomly generate a number between 1 and 237.



→ In this example, the randomly generated number is 26. To identify the corresponding chart, start at the beginning of the hypertension register and count to the 26th entry in the register.

→ Find the corresponding chart and check that the patient meets the inclusion criteria.

- ◆ The patient must have been 18 years or older 12 months before the date of extraction.
- ◆ The patient must have visited the clinic within the past 12 months.
- ◆ If the patient does not meet these two criteria, put the chart back and do not extract data from it.
- ◆ If the patient does meet these two criteria, include the patient and extract their data.

→ Repeat the above steps until the total sample size has been reached.

Worked example: 12-month follow-up

- For example, we are interested in measuring the change in practice following an intervention (such as training health-care workers).
- Assume that the training was completed in March 2019.
- We will therefore use 1 April 2020 as the 12-month cut-off point. This means that patients seen after 1 April 2020 but not in the preceding 12 months do not meet the inclusion criteria. Patients whose last visit was in or before March 2019 do not meet the inclusion criteria.
- Your project’s inclusion criteria may be different, but in some cases patients must have visited the health facility during the inclusion time frame to be eligible for the sample.

Mar19	Apr19	May19	Jun19	Jul19	Aug19	Sep19	Oct19	Nov19	Dec19	Jan20	Feb20	Mar20	Apr20
Train	Intervention time frame (one year) → patient must have been seen during this time frame											collect	

3. Extract the data from individual patient records

Split the records equally among the data extractors or work collaboratively to extract the data, whichever you feel is the most rapid method.

You have two options for recording extracted data.

1. You can also use paper forms that you can print before you go to the clinic. You will need one form per patient. (paper form).
2. You can use a spreadsheet on a laptop, with no Internet connection required (offline form).

Your team may choose to use free products create a computer form to enter data. This can help keep the data more organized and standardized.

Table 1 lists the data extraction questions and variables you will be asked to find in the patient chart and record using whichever method you choose.

Table 1. Data extraction questions and variables

Question	Notes
What is your name?	Name of the person extracting the data
Date of the data extraction (dd/mm/yyyy)	
What is the name of the clinic from which you are extracting data?	
Date of birth (dd/mm/yyyy)	Make sure the patient is at least the minimum age during the intervention time frame to meet the inclusion criteria.
Date of the last visit to the clinic (dd/mm/yyyy)	

Sex (male/female)	
Smoking status (smoker/nonsmoker)	If the patient quit smoking more than 12 months ago, record them as a nonsmoker.
Has this patient been formally diagnosed with hypertension? (yes/no)	
Date of hypertension diagnosis (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data are extracted.
Can you find one or more blood pressure measurements? (yes/no)	
Most recent systolic blood pressure (mmHg)	
Most recent diastolic blood pressure (mmHg)	
Date of most recent blood pressure measurement (dd/mm/yyyy)	
Can you find a second most recent blood pressure measurement? (yes/no)	
Second most recent systolic blood pressure (mmHg)	
Second most recent diastolic blood pressure (mmHg)	
Date of the second most recent systolic blood pressure measurement (dd/mm/yyyy)	If the date is not in the chart, put the date of data extraction.
Diagnosis of diabetes (type 1, type 2 or no)	
Can you find one or more HbA _{1c} measurements? (yes/no)	
What is the most recent HbA _{1c} measurement? (include the units)	
What was the date of the measurement? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.

Can you find another HbA _{1c} measurement? (yes/no)	
What was the second most recent HbA _{1c} measurement? (include the units)	
What was the date of the second most recent HbA _{1c} measurement? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.
Can you find one or more fasting blood glucose measurements? (yes/no)	
What is the most recent fasting blood glucose measurement? (include the units)	
What was the date of the most recent fasting blood glucose measurement? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.
Can you find another fasting blood glucose measurement? (yes/no)	
What was the second most recent fasting blood glucose measurement? (include the units)	
What was the date of the second most recent fasting blood glucose measurement? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.
Can you find one or more total cholesterol measurements? (yes/no)	
What is the most recent total cholesterol measurement? (include the units)	
What is the date of the most recent total cholesterol measurement? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.
Can you find another total cholesterol measurement? (yes/no)	
What is the second most recent total cholesterol measurement? (include the units)	

What is the date of the second most recent total cholesterol measurement? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.
Was the patient prescribed a statin? (yes/no)	
What was the date of the prescription? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.
What was the drug and dose?	
Does the patient have an existing cardiovascular disease? (yes/no)	
Write the existing cardiovascular disease diagnoses (such as angina)	Common cardiovascular disease diagnoses include angina, myocardial infarction, congestive heart failure, peripheral vascular disease, atrial fibrillation and documented cardiovascular or cerebrovascular complications. Check all that apply.
Has the patient been prescribed aspirin? (yes/no)	
What was the most recent date aspirin was prescribed? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.
Has the patient been prescribed a drug for lowering blood pressure (antihypertensive medication)? (yes/no)	
What was the date of the most recent prescription of a drug for lowering blood pressure? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.
Can you find a documented cardiovascular disease risk score? (yes/no)	
Enter the value of the most recent documented cardiovascular disease risk score.	Include the units (usually percentage)
What was the type of cardiovascular disease risk score?	Examples include WHO/ISH, ESC SCORE, Framingham and HEARTS
What was the date the risk score was documented? (dd/mm/yyyy)	If the date is not in the chart, put the date on which the data were extracted.
Can you find a weight, height or body mass index (BMI)? (yes/no)	Answer yes if you can find one or more of the three.

Weight (kg)	
Height (cm)	
Body mass index (BMI)	If you have already recorded weight and height, you do not need to calculate BMI yourself. If the BMI is also recorded, please write this.
Please record any important notes about the data extraction here. Examples include an error you think may have been made, clarification of the units for measurements (such as mmol/L versus mg/dL) or notes you would like for yourself.	

4. Send important information

Keep a record of any of your observations that you think are important to keep in mind when analysing the data. Examples may include any standardized forms the clinic used to capture information, the availability of laboratory testing, the availability of medication and your own observations of the practice, how the records were stored, etc. Share this information with your project coordinator.