

# MEASURES MANAGEMENT SYSTEM

## Measures Testing Tools

Electronic clinical quality measures (eQMs) exist in a fluid environment. They are updated annually not only to address changes in clinical practices, but to address the addition and removal of codes, and to fix bugs and measure logic. This presents considerable challenges because eQMs can be very complex. Prior to the development of testing tools, eQMs were checked manually, which was both time consuming and error prone. In the first year, 381 logic errors were reported, with at least one in each eQM. Since the implementation of testing tools, that number has decreased, but bugs still find their way into eQMs, which is one reason why they are updated on a yearly basis and highlights the importance of having repeatable testing tools within the measure development process.

In addition, new eQMs may be added annually, while others are retired. The standards used to define and report on eQMs also change over time, such as the recent move from Quality Data Model (QDM)-logic based eQMs to Clinical Quality Language (CQL)-logic based eQMs. Due to these types of changes, certified electronic health record (EHR) systems should be tested to ensure they are correctly calculating the updated eQMs. Three current eQM testing tools are discussed.

### Bonnie

Bonnie was first released in 2014 and was used for the annual update processes from 2015 through 2018. Bonnie is a testing tool that enables repeatable automated testing of eQMs which led to a reduction in errors by 85% within the first year. Bonnie provided faster measure testing, as well as greater confidence by CMS and the community at large that the measures were actually calculating in the way that they were *intended* to calculate. Bonnie enables people to load eQMs that have been exported from the Measure Authoring Tool (MAT). This allows developers to:

- Build synthetic patients with history elements that are using those data elements specified within the eQMs
- Test new and updated eQMs using those synthetic patients
- Explore the behavior and complexity characteristics of eQMs

An important metric within Bonnie is logic coverage. This enables measure developers to see and understand how much of the logic they're executing within their test deck. For 2015, the goal for measure developers was 80% logic coverage of measures, but for 2016-2017 and for 2018 the target was to get 100% logic coverage, while ensuring that the eQMs are hitting each piece of logic within Bonnie.

Measure developers can use Bonnie to create *synthetic* patients for each measure (see Figure 1) and can input metadata about the patient, such as:

- First and last name
- When they were born
- Race and ethnicity, gender
- Any descriptions that are needed for what exactly this patient should be testing.

# MEASURES MANAGEMENT SYSTEM

The screenshot shows the BONNIE Patient Builder interface. At the top, there's a navigation bar with 'Dashboard', 'Account', 'Help', and 'Logout'. Below that, a header shows 'Measure Period: 2012', 'CMS104v6', and 'Create new patient'. The main area is divided into several sections: 'TEST PATIENT' with input fields for 'LAST NAME\*' (Smith) and 'FIRST NAME\*' (John); 'PATIENT DESCRIPTION' with a text area; 'DATE OF BIRTH\*' with a date and time picker (8:00 AM); 'LIVING STATUS' with a 'Deceased' checkbox; 'RACE' (American Indian or Alaska Native), 'GENDER' (Male), 'ETHNICITY' (Not Hispanic or Latino), and 'PAYER' (Other) dropdowns. On the right, there's a 'DISCHARGED ON ANTITHROMBOTIC THERAPY' section and an 'EXPECTED VALUE' section with input fields for IPP, DENOM, DENEX, DENEXCEP, and NUMER. At the bottom, there's a 'PATIENT HISTORY' section for 'CMS104V6' showing 'CQL Calculation Results' and an 'INITIAL POPULATION' dropdown.

Figure 1: Bonnie's Patient Builder View

Additionally, developers may add the expected value. These are the expected calculation results, i.e. how this patient is expected to fare according to the conditions defined within the measure. The utility of the testing tool is that if the patient calculates differently than what is expected, Bonnie will show that as a *failed* patient calculation. This likely means that there is an issue within the measure logic which will require the measure developer to go back and address any logic issues discovered through this testing process.

Bonnie uses the Health Level Seven International (HL7) Health Quality Measure Format (HQMF) standard. This is the logic wrapper for the measure logic that holds all the information regarding the measure. In addition, there is the HL7 CQL, which is the logic representation for the measure. The QDM is the data model representation of the measure, which defines what a data element looks like and what attributes are on a data element such as relevant period or author date and time.

Bonnie also supports the HL7 Quality Reporting Document Architecture (QRDA) format. After a test deck has been created and a new publication of the QRDA has come out, Bonnie enables measure developers to *export* patient packages in the QRDA format.

Please visit the [Bonnie website](#) to learn more. Bonnie is an open source development project and the code that goes into creating Bonnie can be found at <https://github.com>.

## Cypress

[Cypress](#) is an open source project freely available for user adoption by the health information technology (IT) community which includes EHR vendors as well as testing apps. Cypress is a testing tool that clarifies or validates whether an EHR system can correctly calculate eQMs.

Given a set of test patients and eQMs, Cypress tests the ability of an EHR to:

- Capture the information required to calculate the selected eQm.
- Export the information required to calculate eQMs
- Calculate the results of selected eQMs

# MEASURES MANAGEMENT SYSTEM

- Report the calculated results of the selected eCQMs (this would include reporting the eCQMs to CMS and so complying with the CMS QRDA standards and implementation guides)
- Report filtered eCQM data

One of the key components of testing in Cypress is a set of test patients that is provided to the vendor system to enter into their EHR and is generated using the Bonnie tool. The 2016 test deck included 93 measures which had 81 eligible clinician patients and 81 eligible hospital patients. In that bundle Cypress had a minimum coverage of 80% of the logic for all those measures.

Table 1: Cypress Bundles 2016 and 2017

	2016	2017
# of Measures	93	71
Eligible Clinicians Patients	81	205
Eligible Hospitals Patients	81	78
Minimum Coverage	80%	95%
Populations	Every patient population covered by test deck	

The bundle released in 2017 was for 71 measures and included just under 300 patients. Cypress increased minimum coverage from 80% to 95% coverage. Many of those measures had 100% coverage. Every patient population is covered in the test deck, and so that means that every population-defining eCQM has a patient that corresponds with that logic. See Table 1 for a comparison of 2016 bundles with 2017 bundles.

Cypress is made available as an open source tool which means that the source code is made freely available and can be redistributed as well as modified. There are also released versions of Cypress that include a Virtual Machine, which is a fully enclosed version that can be installed as-is on an organization's network.

Cypress also has [public issue tracking](#), which allows the user community to report bugs and issues. Prior to major releases, Cypress releases beta versions seeking input from the user community for usability as well as finding bugs. The next release is scheduled for July 2018, and beta versions will be available preceding that release. Most of the usability enhancements incorporated into Cypress have been user-driven — users getting in contact with us from the Google group as well as the JIRA site. There are regularly scheduled Cypress Tech Talks open to people interesting in learning more about Cypress. See the [Electronic Clinical Quality Improvement Resource Center](#) for the schedule and to register.

## National Committee for Quality Assurance (NCQA)

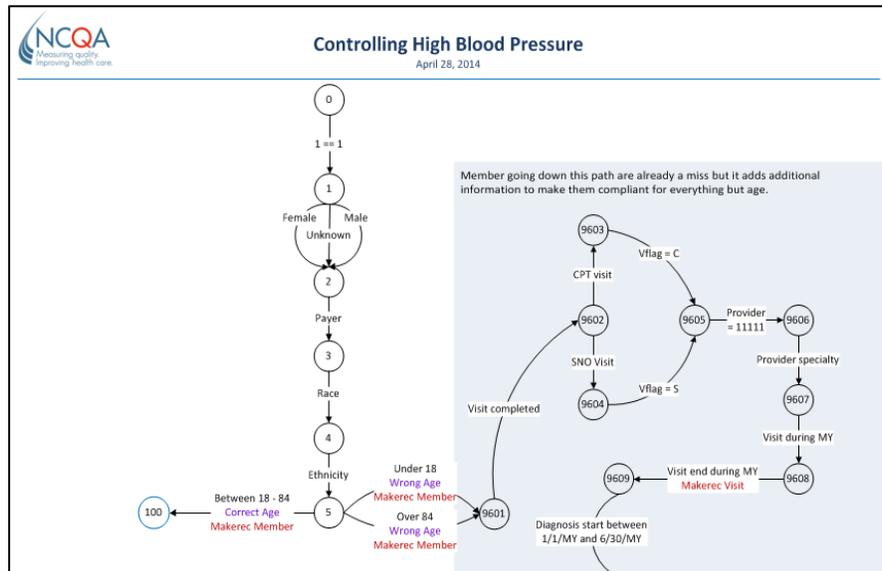
The NCQA eCQM test deck development process generates synthetic test patients by measure and then looks at that measure to examine all the decision points within that measure. The NCQA process also considers common programming errors that could happen with that specific measure such as:

- Where have measure developers struggled to express desired clinical conditions in the logic?
- Where is the logic unclear, complicated, or difficult to interpret?

The NCQA process aggregates all these points and creates a flow diagram for the measure. This flow diagram (Figure 2) captures all the decision points within the measure as well as decision points to be tested to make sure that people have the measure appropriately calculated. For instance, a decision point for sex — either male or female — will be generated even if the measure requires both males and females, to ensure that a vendor is not inadvertently excluding either males or females from the measure.

# MEASURES MANAGEMENT SYSTEM

Figure 2: NCQA Process Flow Diagram



Once the flow diagram is generated, three programs are run. The first program ensures that every patient goes through the flow diagram on a different path, and that all possible paths are hit by at least one patient within the set of test data. The second program then takes those patients and follows their path through the flow diagram and generates the patient data. For instance, if a patient is flowing down a path that says he is aged 18-85, then the second program will generate somebody who is 18-85. The first run-through may generate somebody who is 56; the second run-through may generate somebody who is 25, but that birthdate will fit within the parameter specified within the flow diagram. The third program formats the data to send to the vendors, using the standard formats provided by HL7 to send patient data to the vendors. This allows the process to generate many sets of test data. Each set of test data for a measure will look completely different.

In the first set of test data generated, the first person in that data may be a numerator hit. They will be the right age and have the right services for the numerator. The second time the program runs things will generate in a different order, and the first patient will be a miss. The patient will not even be in the measure denominator, but the process can generate as many sets of test data desired, anywhere from 800-1,500 patients per measure to make sure that all the possible paths are tested through that measure — as well as testing as many of the codes and value sets as possible for each of the measures.

Once the vendor has generated their result files, they upload them on the NCQA website to be scored immediately. That scoring report will return just a simple pass if all is agreed. If there is disagreement on some things and the match is not 100%, then the vendor can look at their data.

If vendors would like to test with the NCQA, please visit [the NCQA ONC health IT testing website](#) on the NCQA family of websites. Getting set up with NCQA is very easy. There is an application form to help NCQA identify how many measures the vendor wants to test and what version of the specifications to test.