



**D3 – Designs, Data, Decisions**

**“Is Our Survey Data Good Enough for Public Policymakers?  
Field Quality Control Techniques in F2F and CATI Surveys at D3  
(1987-2017)”**

**Presented at the 70<sup>th</sup> Annual Congress of the World Association of Public  
Opinion Research, July 15-17, 2017, Lisbon, Portugal.**

Dr. David A. Jodice, Founder & CEO, Mr. David Peng, Chief Statistician,  
David Rae, Statistician - Statistical Programmer, Timothy Van Blarcom, Statistician –  
Survey Methodologist, and Angela Ulrich, Assistant Statistician

**Summary**

For more than 30 years D3 and its principals have been implementing an evolving mix of quality-assurance techniques across the thousands of surveys we have conducted in over 120 countries. This paper focuses on our efforts to prevent data collection errors (such as outright fabrication and incorrect technique) one of the key elements within the Total Survey Error framework. D3 is increasingly conducting monitoring & evaluation studies, the results of which are often directly linked to public policy, in terms of moving programs forward or suspending them. So, are our survey data good enough to inform public policy makers on the effectiveness of their choices as well as reliably track public opinion on a wide variety of issues?<sup>1</sup>

---

<sup>1</sup> In addition to process validity which we focus on here, D3 has an ongoing effort to assess field quality through outcome validity testing – the comparison of survey results with election outcomes, in countries with fair and free elections. However, as elections are episodic and not always fair and free, process validity testing will remain our key method for assuring quality.

We have worked our way through six forms of field quality assurance over the past 30 years, all of which has some applicability today.

- Buying by the Brand of the Local Practitioners,
- On-Site Observation of Data Collection,
- Employing 3rd Party Evaluators of the Field Work,
- Statistical Analysis of the Full Body of Work,,
- Cross-Border Polling from Regional Call Centers, and
- Conducting F2F Surveys with Mobile Devices, Using our Proprietary Software

### **Total Survey Error – Setting the Framework for Quality Assurance**

“Total Survey Error (TSE) refers to the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data. A survey error is defined as the deviation of a survey response from its underlying true value.”<sup>2</sup>

Key elements in total survey error fall into two categories: Errors of Representation and Errors of Measurement. Representation issues arise from coverage, sampling, non-response and adjustment errors. Measurement issues arise from construct validity, measurement in the field and processing errors. Another key source of error arises from inferential error – the interpretation and application of the findings by the client and/or end user.<sup>3</sup> Errors may also arise in the design phase, by selecting an approach that is not appropriate, failing to build in agreed-upon benchmarks and not balancing the research strategy with resources.

### **Focus on a Key Element of Measurement Error: Interviewer Effects**

For D3, a company best known for its data collection efforts, particularly in hard-to-reach conflict and post-conflict environments, field work is our key focus. For D3, that means the right sampling frame and fully transparent execution of the sample and properly conducted field work.

In this paper we are focusing on our history of managing field work outside of the United States, including setting standards, participating in briefings of the team, observing field

---

<sup>2</sup>For extended treatments of TSE, see Robert M. Groves and Lars Lyberg, “Total Survey Error: Past, Present and Future,” and Paul P. Biemer, “Total Survey Error: Design, Implementation, and Evaluation,” both article are from *Public Opinion Quarterly*, (Volume 74 Number 5, Special Issue, 2010); Brady T. West and Annelies G. Blom, “Explaining Interviewer Effects: A Research Synthesis,” *Journal of Survey Statistics and Methodology* (Volume 5, Number 2, June 2017), pp. 175-211.

<sup>3</sup> For more on Total Survey Error, see Darren Pennay, “Total Survey Error,” a QPMR Webinar of March 17, 2014.

work and back-checks and using a variety of statistical tools to analyze each interviewer's full body of work.

The interviewer effects on survey data quality fall into two categories: (1) errors and omissions and (2) malfeasance or outright cheating. Neither of these are new concerns nor are limited to developing countries or conflict environments. Not surprisingly, as the country which created scientific opinion research, published documents about data quality reach back to the 1940s.<sup>4</sup>

### **The Role of the Client**

The client has a key, but often overlooked role in quality assurance. Clients can improve the likelihood of a high quality (with valid and reliable data) survey in several ways:

- Matching requirements to the budget in the design phase,
- Clearly setting expectations and benchmarking standards at the outset,
- Allowing the agreed-upon time frame for field work and back-checking,
- Participating in field briefings and observing field (when practical), and
- Properly supporting the role of independent, third-party evaluators (when desired).

### **Buying by the Brand of the Local Practitioners**

In the late 1980s, D3 began with a brand-based approach to vendor selection and quality control – we would buy research services from well-established and highly respected companies and rely on them to take care of quality control in their spheres of sampling, interviewing and data processing. We would focus on design and analysis.

This was a typical practice for vendor selection across American businesses. As much of our early work was in Western Europe, this strategy was very successful with such market leaders as the Emnid Institute (Germany), Ifop (France), NIPO (Netherlands) and Doxa (Italy), among many others. Our early contacts with the Gallup International Association (not associated with Gallup, Inc.) provided us with access to new colleagues well versed in the methodology of public opinion polling.<sup>5</sup>

A key element of brand-based purchasing is the use of outcome validity testing. For opinion pollsters, the quality of their work can be assessed in terms of survey findings published

---

<sup>4</sup> See for example, Edward Demings, "On Errors in Surveys," *American Sociological Review*, 1944, (9), 359-64.

<sup>5</sup> With the growth of large multinational research chains and organizations, the Fortune 500 buyers of regional and global market research surveys are now buying by brand as a standard corporate practice – putting out RFPs to Kantar, Gfk, ISPOS, Gallup Inc., and others.

before an election and then compared with actual election results. Of course, for this to work we need fair and free elections and valid vote counting.

### **On-Site Observation of Data Collection**

We soon moved beyond Western Europe and experienced greater challenges in conflict and post-conflict zones as well as emerging markets. Brand was no longer an appropriate guide to vendor selection or assurance of quality. Working with new firms, some of our own creation, (MIB in Bosnia & Herzegovina since 1996 and ACSOR Surveys in Afghanistan since 2003), we supplemented observation of training and field work with the statistical analysis of the whole body of work of each interviewer in each survey. We have also welcomed on-site client participation and have funded 3<sup>rd</sup> party evaluators for on-site visits to the F2F field operations (sometimes at our own initiative, sometimes as a client requirement).

On-site participation in the training of the field team and observation of field work (household and respondent selection, interviewing and back-checking) are among the true joys of international public opinion research. Over a long period of time Dr. Jodice has been out in the field in Bosnia & Herzegovina, China, Dominica, Egypt, Russia, France, Germany, Grenada, Italy, Jamaica, Kazakhstan, Kosovo, Macedonia, Mexico, the Philippines, South Africa, Tanzania, Trinidad & Tobago, Turkey, and most recently, India. Other D3 staffers regularly travel to observe field. Not only do we benefit from seeing how field is actually done and can thereby evaluate interviewers and improve training, we also benefit from an increased understanding of the challenges which our international field teams face. In a number of cases, we have been able to include clients in these field observations and that has provided many benefits to all parties.

On-site observation of field has one problem – it's not scalable. With samples ranging from 1,000 to 5,000 or more, D3 and its staff are not able to observe a significant share of the field work. And, unless one selects interviewers to follow based on previous performance, the local field team can be expected to send us out with their best supervisors and interviewers.

### **Employing 3<sup>rd</sup> Party Evaluators of the Field Work**

In a number of countries, D3 has supplemented our own staff with third-party evaluators. Across the Balkans in the early 1990s, we contracted with well-established researchers from EU and Middle Eastern countries to observe fieldwork in Bosnia, Kosovo and Macedonia, not only to leverage our staff but to bring an independent perspective to these relatively new field operations. In other countries (Afghanistan in particular) clients have contracted with third party researchers to evaluate our work and report directly back to the client. This is often a requirement of the stakeholders/end users/funding organizations and we have no objection to it in principle. Of course, if the client is able to send their own staff,

so much the better, but this is often not possible in conflict environments such as Afghanistan and Iraq.

The key problem we have experienced with third party evaluators selected by the clients is the presence of an inherent *institutional conflict of interest*. It is rarely the case that the evaluators are not technically qualified. However, the firms or individuals selected to evaluate our field performance have been, in a disturbing number of instances, those who did not win the project after a competitive solicitation. This raises two issues:

- To what extent should D3 and its field team (either a subsidiary or an independent firm) be required to release proprietary information on sampling and field practices to a competitor, under the auspices of a methodological evaluation? And,
- How can a third party with such an institutional (or personal) conflict of interest make an independent and unbiased assessment of the performance of the D3 or local field team?

This problem originates with the client, but it can affect all of us, with a serious impact on both the business and methodological sides of our work. Clients across the board (but in our experience this has occurred more with national and international public sector organizations and their prime contractors) need to build into their practices staff training and behavioral monitoring within their organizations to ensure that the technically qualified evaluators they select are free of personal or institutional conflict of interest.

### **Statistical Analysis of the Full Body of Work**

Beginning almost 20 years ago for our F2F surveys, D3 began to statistically analyze the entire results of each of our surveys, focusing on response patterns by interviewer. This solves the scalability problem and reduces the probability of assessment error – as we are able to look at the entire body of work of each interviewer taken individually and as a team under specific supervisors. In a number of countries we were conducting trend surveys (typically monthly or quarterly), so we are able to monitor interviewer performance across multiple surveys. We have been looking for evidence of both unintentional errors by the interviewer and deliberate cheating.

To assess possible field measurement errors, we are testing the data across five indicators that may, together or separately, indicate error or malfeasance:

- Duplicate Interviews,
- Pattern Responses (Straights) Across Questions within Respondents,
- Pattern Responses (Straights) Across Respondents,

- Non-Response Patterns (defined as critical if more than one-fourth of the completed interviews have more than 25% refused or don't know), and
- Field Productivity (average interview length compared to the survey overall and number of completed interviews per working day).

While there are many points which can be made about these indicators, we think the following are most important:

- Duplicate interviews may be defined in terms of having 85%, 90% or 95% (for example) identical responses to the substantive questions. We would expect any interviewer who is duplicating questionnaires to not have 100% identical responses, so we set the bar lower for defining what a duplicate interview is. We see duplicates as a greater concern when they are grouped within the work of one interviewer or under one supervisor. Duplicates are much less of a concern if they occur across the work of different interviewers in different regions of the country.
- Straights across questions or across respondents may reflect the homogeneous nature of the sampling points (rural residency combined with ethnic similarity within the SP) and is more likely to occur with smaller clusters (5 vs. 10 per cluster).
- Non-response patterns may indicate poor performance by the interviewer, cheating or respondent characteristics (such gender, age group, residence and education) which drive their ability to answer the questions. Again respondent characteristics may be concentrated within clusters. Those which are more rural, less educated, more female and older in population tending to have a higher non-response rate in multi-topic public opinion polls.
- An interviewer that is too productive (short interviews, too many completes per working day, too few callbacks, etc.) may be extremely efficient or he/she may be guilty of data fabrication.

Based on our analysis of their full body of work, interviewers are scored into four categories:

- **Commend** – These interviews are not cited in any of the first four indicators and only one of the two productivity indicators.
- **No Action** – These interviews are accepted – The interviewer is not flagged for any of the top four metrics, but did receive flags on both of the two productivity indicators.

- **Yellow** – It's very easy to be coded as Yellow (needing more training, supervision). The interviewer needs to be flagged on only one of the metrics to be coded yellow.
- **Red** - This interviewer has more than ¼ of his/her respondents with 50% non-response, has been cited for any one of the metrics for a yellow or has multiple duplicate cases and is suspected of data fabrication. While D3 cannot fire the interviewers who work as employees of independent firms, we do require that they not work on future D3 projects. <sup>6</sup>

D3 and its suppliers/subsidiaries have a responsibility to improve the likelihood that interviewers will not cheat. We do that in the following ways:

- We clearly indicate our standards for sampling, questionnaire administration and interaction with the respondent.
- We deploy supervisors to observe and back-check interviews on-site.
- We also back-check interviews by telephone from the firm's main office, breaking a possible link of complicity between supervisors and interviewers.
- We conduct the QC assessment for our surveys, but do not explain the metrics to the field team.
- We allow sufficient time and travel funds to conduct the interviews properly.
- Wherever possible we pay field by time or a blend of time and interviews completed. We recognize that paying only by completions creates a moral hazard for the interviewers, particularly with hard to reach populations and remote sampling points
- We insist that field companies use consistent ID numbers for their interviewers and supervisors to enable us to track performance over time.

### **Cross-Border Polling from Regional Call Centers**

---

<sup>6</sup> With client approval, D3 will eliminate duplicate cases from the final dataset. We often, again depending on budgets and client requirements, will oversample to allow for the deletion of interviews which are completed by interviewers who are rated Red. In some cases, we may deliver data files with and without the questionable cases, allowing clients to see the data both ways.

With the expansion of violence across the Middle East and the growth of mobile phone coverage across the region, D3 and its suppliers are increasingly using CATI to reach large-samples of the general population. For many projects it makes sense (cost efficiencies, safety for the field team and respondents, greater freedom for questionnaire content and methodological consistency) to conduct national and multi-country surveys from a regional and multi-lingual CATI facility.

While we lose some of the contextual feel provided by F2F interviews (particularly understanding the household's socio-economic status), the gains of cross-border polling far outweigh those losses. The gains include:

- Safety of the field team from violence on the ground in the survey country,
- Freedom of the survey process from local censorship,
- Improvement of quality assurance through real-time observation of interviews by the call center's supervisory team,
- Recording (when permitted) of actual interviews for evaluation and training purposes,
- Sampling through the use of RDD-based replicates of telephone numbers (generally dual-mode, but increasingly tending toward mobile dominant samples), and
- Tracking each interview by start and end times and the number of seconds taken by each question. This is improved quality control over the field team.

### **Conducting F2F Surveys with Mobile Devices, Using our Proprietary Software**

D3 is increasingly replacing our long-term F2F quality assurance methods with CAPI tools. While overall survey cost and work pace are important, this transition is also driven by the proliferation of lower cost mobile devices, increasing network accessibility within large numbers of emerging markets and rising client requirements regarding sampling, quality assurance, survey administration and contact disposition reporting. Our approach is based on Android-compatible devices operating our proprietary software: Research Control Solutions. RCS is more than a data capture tool. It enables us to independently track location, interview length and time between interviews, as well as providing real-time (or end of the day) uploads of survey results. It also offers customizable dashboards for staff, field suppliers and clients to observe field progress and results.

The quality assurance aspects of these new tools benefit D3 and its clients in several ways:

- Automatic capture of geo-location data through GPS (more often for the entire sampling point, not the participating households, unless they give permission),
- Automatic date and start/end time capture (for the contact and the actual interview),
- Tracking time by item throughout the interview,
- Recording of all or portions of the screening process and the interview (with the respondent's permission),
- Photos linked to the interview results of the sampling point and/or with respondent permission their household (often not desirable or permitted),
- Automatic data capture and real-time (or later in the day) uploading when WIFI or 3G/4G network access is available,
- Automation of key questionnaire components (filtering, response rotation and randomization), and
- Real-time recording of contact disposition codes

While there are many benefits for survey administration and documentation, the key factor that is driving client interest in mobile based CAPI techniques is the automatic capture of data that proves **When** and **Where** the research was conducted. This takes us right back, in a very positive way, to quality assurance concerns about data fabrication and solutions.

## **Where Are We Headed?**

The prior sections of this report describe various areas of quality assurance that D3 has developed, invested in, and implemented throughout its history. In this section, we provide an overview of where D3's most recent research and development in QC is headed.

### *A. Continuous Process Improvement*

D3's goal of quality control is not simply to remove potentially erroneous data from final survey datasets prior to analysis and/or client delivery. While this is the most immediate use of our quality control analysis, it is not the primary long term goal. Instead, for multi-study projects or long-term field partnerships, our goal is to improve the entire process by identifying specific issues that may have resulted in questionable survey quality and implementing improvements where necessary (interviewer training, supervisor training, data processing, incorporation of new technology such as tablets, etc.), all with the ultimate

goal of reducing the number surveys that may have questionable quality, thus improving the overall product.

This is analogous to quality control practiced in the fields of engineering and manufacturing, where continuous process improvement<sup>7</sup> is stressed. We ultimately do not want to spend time, effort and money producing but then throwing out low quality products, we want to stop producing them all together or at least keep them to a minimum.

In order to implement continuous improvement in the field of survey research, D3 project managers and analysts must be able to spend time exploring the data and working with field partner staff to identify the potential source(s) that require special attention and/or retraining. However, given the modern-day quick turnaround expected for survey data, there is less and less time to spend properly exploring QC testing results because so much time is spent performing the actual testing itself. This promotes the immediate remedy of removing cases that could bias the overall survey results due to quality issues, but does not help with long term continuous process improvement.

Therefore, D3's Statistics and Analytics team recently set into motion the development of an interactive dashboard that performs the aforementioned *Statistical Analysis of an Interviewer's Full Body of Work* in a simple point and click interface. The dashboard then populates various versions of an interactive report that allow the analyst to investigate the data without spending too much time performing the actual QC testing.

### B. valkyRie

The valkyRie project is the result of this long-term goal: a point-and-click, easy to use, semi-automated<sup>8</sup> QC system for project managers and analysts. In a nutshell, valkyRie has 4 primary components:

1. A user interface (dashboard) that is launched through an internal website. This allows users (D3 project managers and analysts) to upload datasets, select variables, and download reports using a point-and-click interface.
2. In the background, code has been written in the R computing environment to interpret those user inputs with the uploaded data, and perform the appropriate testing on selected variables.

---

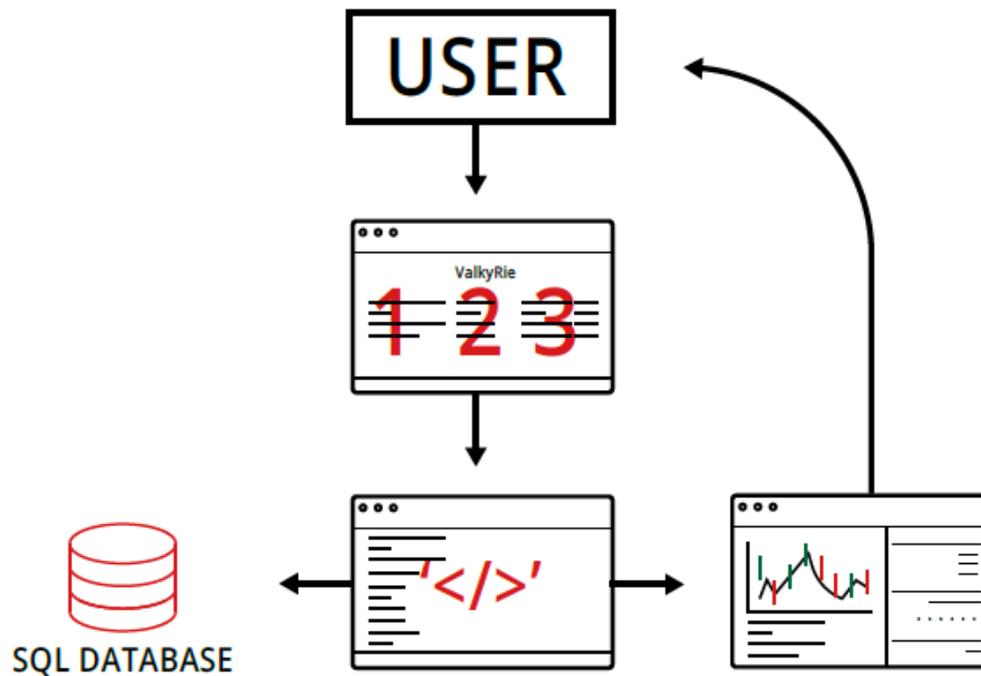
<sup>7</sup> <http://asq.org/learn-about-quality/continuous-improvement/overview/overview.html>

<sup>8</sup> We say it is semi-automated because we still require user-input on variable identification and data upload. Fully automated systems are possible, but may result in users being too far removed from understanding what they are doing.

3. The results of these functions in Rare then stored in a SQL database.
4. A dynamic HTML report is published with these results that allows users to explore cases that warrant more attention through some interactive graphs. The report also provides recommendations based on the results.

As Figure 1 indicates below, all of these capabilities of valkyRie center around the R Computing Environment<sup>9</sup>, which not only runs each of the tests that we will introduce in the next section, but also handles the interaction between the graphical user interface (GUI) dashboard, as well as the storage of results into a SQL database. The latter provides us with an exciting capability to expand analysis beyond a single survey, which we will discuss in a subsequent section.

Figure 1: Flowchart of valkyRie



---

<sup>9</sup> We must give credit to the authors of packages such as *Shiny*, *tidyverse*, *Rmarkdown* as they provided us with the tools to create valkyRie.

To the user, the whole process is handled via a website. The user enters the address (which is currently internal to D3's network) and comes to a website.

Figure 2: Screenshot of valkyrie Prototype Dashboard

**D3 designs • data • decisions**

### Valkyrie Quality Control Tool

Process Document - Z:\Team E\Resources\Process Documents

**Step 1: Survey Identification**

Select the Vendor: ACSOR

Select Project: ANQAR

Enter Wave Number: [Text Field]

Country: Afghanistan

Select Mode: CATI

**Step 2: Upload Dataset**

Upload a SAV which you wish to conduct a QC Analysis on

Upload a SAV: [Browse... No file selected]

Download a .csv in order to specify substantive and battery variables

Download List of Variables

Upload downloaded variable list .csv with any changes: [Browse... No file selected]

**Step 3: Update Variable Names**

Enter the following variables exactly as they appear in your sav, case sensitive

Interview ID: m12

Interview Year: m8

Interview Month: m9

Interview Date: m10

Respondent ID: m1

Length of Interview: m17

Supervisor: m14

Province: m7

Run Data Cleaning Tests | Download Results | Run QC tests | Download QC Report

Here, the user can upload a survey dataset. He/she then simply specifies which variables should be included in the QC test (at this point, they should be familiar with the questionnaire), and the tests are run automatically. A dynamic report is then populated with the results of the QC, with several interactive graphs embedded within the report that allow the user to do some preliminary exploration.

There are numerous benefits to this semi-automated process:

1. Reducing analyst time needed to perform QC testing and preparing a report by an estimate of more than several hours. This time is freed up to spend investigating errors and working on solutions with our field partners.
2. Consistency
  - a. Although the analysts still have input on which variables are to be used, they can be assured that all of the QC tests will be conducted the same way because they are automated.
  - b. Consistency in formatting and reporting. All of the reports will be auto populated with key test information and outcomes.

- c. Consistency in performing the QC for all projects. Given the reduced amount of staff time required to at least run the QC, it should be executed for all quantitative projects, whether explicitly required by the client or not.

### 3. Dynamic Reporting

The output report is dynamic. Once the dataset is uploaded, and the user inputs and makes some basic selections of variables, the report is automatically populated.

However, we have opted for a dynamic report that is an HTML output. This allows us to leverage some of the interactive graphs to allow for users to look at the data while reading the report.

### 4. Database Development

The most powerful aspect of valkyrie is the integration of QC results across all D3 projects. This ever-growing database will allow for comparisons beyond just the individual project, but by field partner, country, region, type of survey, etc.

We expand below on the latter two points, dynamic reporting and building a growing database of quality control.

#### C. *Dynamic Reporting*

Static reports (Word documents, PDFs, Powerpoint) remain the most common outputs and document types to share QC results through.

However, D3 is prototyping dynamic reporting via HTML reports produced using *R Markdown*. We show a few example screenshots of the features we have developed thus far.

#### 1. Side panels

We take advantage of packages authors have created in *R Markdown* to simulate the Edward Tufte side panel usage.

Figure 3: D3 Dynamic Interviewer Quality Control Report Prototype

**Interviewer Quality Control Report**  
*D3 - Design Data Decisions*  
 May 22, 2017

**Test 1: Duplicate Cases**

The first test is a duplicates test. The D3 project manager identifies a list of variables in the dataset from which to perform this test, excluding certain variables that we expect a high number of duplicates (ex: filtered questions). Our program then cycles through to compare each interview, one-by-one, with other interviews. Therefore, all combinations of interviews are compared. Pairs of interviews that had the same value recorded in at least 90% of the variables analyzed are displayed in the table below.

**Table 1: Duplicates Table**

Percent Similar (>= 90%)	ID		Interviewer		Province		Supervisor	
	Case A	Case B	A	B	A	B	A	B
0.907	113	719	15	4	7	12	3	5
0.948	115	799	15	3	7	12	3	5
0.918	203	1378	4	1	12	12	5	5
0.918	425	723	4	2	12	12	5	5

While these are candidates for deletion (if the contracted sample size allows), D3 Statistics and Analytics recommends focusing on improving the overall process improvement as the ultimate goal of QC (as opposed to just *cleaning* data). This requires investigation by project managers and partner firms. As a result, this report contains some additional graphical input to help analysts think about the flagged cases from different points of view.

Figure 1: Boxplot of Interviews above 90% similar, by Interviewer

D3 checks for all pairs of interviews that are above 90% identical, particularly those pairs where both are conducted by the same interviewer, Interviewer, or pairs of interviewers, responsible for multiple duplicate interviews will be flagged by this test.

The center line of the boxes show the median (midpoint of the 90%+ similar interviews observed for each of these interviewers).

## 2. Scrollable data/tables

Furthermore, tables often contain too much data and lead to lengthy reports. We made scrollable tables within the report that allow a user to scroll down within a small segment in the window:

Figure 4: Scrollable Table Example

**Table 1: Duplicates Table**

Percent Similar (>= 90%)	ID		Interviewer		Province		Supervisor	
	Case A	Case B	A	B	A	B	A	B
0.907	113	719	15	4	7	12	3	5
0.948	115	799	15	3	7	12	3	5
0.918	203	1378	4	1	12	12	5	5
0.918	425	723	4	2	12	12	5	5

This avoids unnecessarily lengthy static reports and/or the need to flip back and forth between two different types of files (ex: Excel and Word) or having to go to an Appendix.

Figure 5: Scrollable Table Example Continued

Table 1: Duplicates Table

0.938	463	828	14	18	3	5	2	2
0.907	464	829	15	19	3	5	2	2
0.918	465	823	16	19	3	5	2	2
0.918	466	824	13	20	3	5	2	2
0.938	467	825	14	17	3	5	2	2
0.938	468	826	15	20	3	5	2	2

By leveraging the side panel for links and downloads, a user can still obtain the raw data in their desired format for more in-depth analysis.

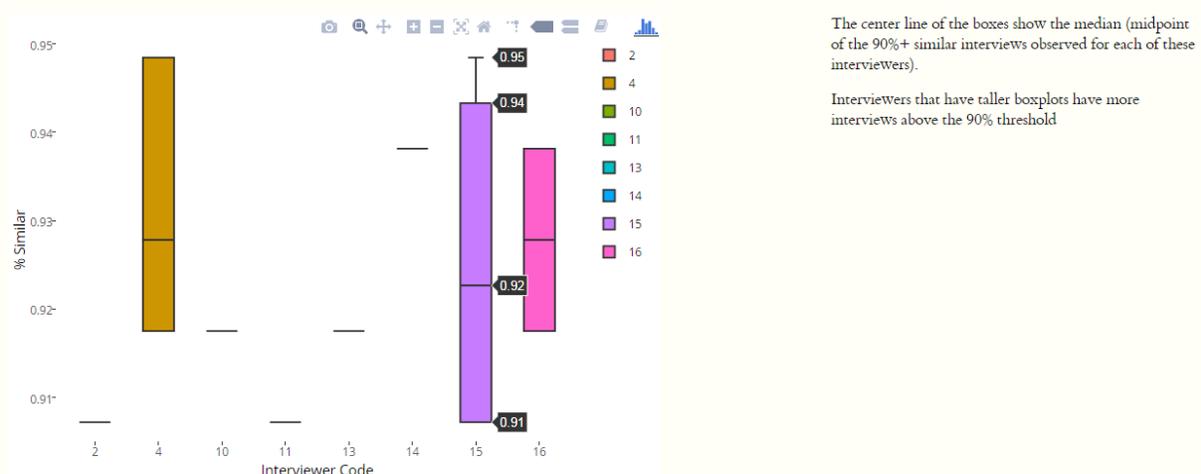
### 3. Powerful and Interactive Graphs

Furthermore, *R Markdown* in combination with *Shiny*, *ggplot2*, and *plotly* allows us to embed interactive graphics. As you can see in Figure X below, the data points will show once the user scrolls over these boxplots.

This allows us to produce graphs that are clean, yet contain additional information. Furthermore, and perhaps most important, is that it promotes engagement with the reader of the report.

Figure 6: Interactive Box Plot Example

Figure 1: Boxplot of Interviews above 90% similar, by Interviewer



#### D. Database Development

In the developing of automating our QC Report, a primary goal was to set up a system that not only produces a report, but also allows us to store results of each analysis into a database from which we can potentially improve the QC process.

As a result, D3 has set up a SQL database using *SQLite*, where results from our tests conducted using R code are stored. These results are also tagged with information that the user initially inputs into the dashboard when uploading the dataset.

For example, users input the country, field-partner and mode:

Figure 7: User Inputted Data Tags

### Step 1: Survey Identification

The screenshot shows a form titled "Step 1: Survey Identification" with the following fields:

- Select the Vendor:** A dropdown menu with "ACSOR" selected.
- Select Project:** A dropdown menu with "The Asia Foundation" selected.
- Enter Wave Number:** An empty text input field.
- Country:** A dropdown menu with "Afghanistan" selected.
- Select Mode:** A dropdown menu with "Face-to-Face (Paper)" selected. The dropdown is open, showing three options: "CATI", "Face-to-Face (Paper)", and "Face-to-Face (CAPI)".

This information is then also stored with the actual QC results allowing for comparisons and expanded analyses across partner/vendor, project, country/region and mode.

Our goal is to expand the database using past, present and future datasets. This database will be expanded each time an analyst uploads a dataset for quality control testing, thus allowing us to consider more data each time a quality control analysis is performed.

## **Conclusion**

As covered in this paper, for more than 30 years, D3 and its principals have been implementing an evolving range of quality assurance techniques in the thousands of surveys we have conducted in over 120 countries.

These techniques include, Buying by the Brand of the Local Practitioners, On-Site Observation of Data Collection, Employing 3rd Party Evaluators of the Field Work, Statistical Analysis of the Full Body of Work, Cross-Border Polling from Regional Call Centers, and Conducting F2F Surveys with Mobile Devices, Using our Proprietary Software

We continue to be heavily invested in quality assurance, and as technological advances permit, are looking for new and innovative ways to improve quality. Our most recent development is a database storing statistical analyses coupled with dashboards that take advantage of SQL data storage and the R computing environment's statistical analysis power.