

CHAPTER SIX



Basic research – skills and tools

Learning objectives

By the time you have worked through this chapter, you should be able to:

- Understand the meanings of 'following a paper trail' and 'data-mining' and apply these techniques in your own country's context, however information-poor
- Have a basic grasp of computer-assisted reporting (CAR)
- Understand the principles of database-management tools
- Use Access to Information procedures to inform and strengthen your investigative stories
- Identify when a story requires figures as well as facts
- Identify what type of numerical information is required
- Carry out basic mathematical operations with numbers
- Understand how statistics are gathered and compiled
- Understand what statistics can (and cannot) convey and how they may mislead
- Identify story ideas from numerical and statistical information
- Ask essential questions about and conduct essential checks on numerical information provided by others
- Develop relevant interview questions from numerical and statistical information
- Convey numerical and statistical information clearly and accurately to readers.

NOTE: This chapter collates and develops themes and skills dealt with in earlier chapters of the IJ handbook. We therefore recommend you read Chapters 1-5 first – and, because of the important ethical aspects of investigative work raised here, also read Chapter 8 on Law and Ethics.

 Konrad
Adenauer
Stiftung



 Fair
FORUM for
AFRICAN
INVESTIGATIVE
REPORTERS

 cijj

The nuts and bolts of investigation

Investigative journalism has to be founded on an understanding of how systems work – or are supposed to work. So, you will ask questions such as

- ❓ How is this process or system supposed to work?
- ❓ Who is supposed to do what, and when and how?
- ❓ What documentation is supposed to exist to record and track the system?
- ❓ What standards are supposed to be in place, how were they established and who enforces them?

The more comprehensive the answers you can build up to these questions, the more likely it is you will build up a picture of the points where things could go wrong – and identify exactly where they did go wrong.

Mapping any area of knowledge has two aspects: qualitative and quantitative. Qualitative mapping is about people, events, reasons, motivations, feelings and arguments. Quantitative approaches put the numbers onto the map: how many quality checks does a medicine need to undergo; how much pollution is in the dam; what have been the trends in city crime over the past five years? Very often, it is the figures that can turn a small local story into a major national investigation, by providing hard evidence that, for example, school dropout figures in your community are typical of a problem that is affecting the whole country.

This chapter introduces you to the basic tools and approaches for investigating number-based information and introduces you to some of the key concepts you will encounter when dealing with these kinds of information.

Figuring out where things go wrong

In 2008, the US government estimated the cost of the Iraq War so far as US\$600 billion. Economist Joseph Stiglitz and investigative writer Linda Bilmes have written a book arguing that the real cost of the war will be closer to US\$3 trillion. Some of this argument is based on different predictions, and different ways of categorising available information: the kinds of things economists are likely to continue arguing about. But, the core of the argument is based on their detailed knowledge of how the war funding system actually works: where money comes from and the way official budget presentations don't show everything. For example, US Defense Department accounting is done on a 'cash' basis; unlike most other types of business accounting, it does not include future spending commitments in its presentations. Stiglitz and Bilmes could not have completed their investigation without this detailed knowledge of the system.

Databases

Basic research – skills and tools: building your own database

In many places in Africa, it is impossible to access written records simply because there aren't any. The municipal office may have some papers stacked in brown paper in a cupboard, but even the mayor probably does not know what they are. There is a phonebook for the region, but it dates from 1993 and nobody seems to know where it is. The local police or jail authorities don't keep a precise log of who they arrest, for what, or for how long they detain them.

There are two ways to investigate issues in such a document-poor environment:

- Your own observations
- Structured interviews of others with relevant memories and observations

Firstly, everything begins with your own observation: **seeing** and **experiencing** something (often when you are conducting consumer research, for example, sitting as a patient in a hospital waiting room, or working undercover as we have previously described).

But after you have experienced, seen or heard something that you define as a priority issue, you need to go the second way: ask other people.

Doing structured interviews

This does not merely mean talking to people in the normal journalistic interviewing style – although you will have those kinds of conversations too. But in addition, you must develop a systematic process for building your own databases and statistics, derived from what people themselves have experienced or witnessed. You need to use structured interviews – that is, compile a list of the same questions that you will ask all your interviewees (although you'll also add extra questions flexibly if something new comes up in an individual conversation). Precisely because there are no written records against which you can compare the oral statements it is vital that you compile information that is statistics-like in nature. You're actually doing a mini-survey.

- 1 Compile a comprehensive list of questions that can establish likely facts – for example, asking everyone if they can remember when something first happened. In this way, you can assess when a certain problem started (rapes and pillages by strangers; crops dying; people with machines conducting excavations; road deterioration; disappearances of local people); its possible causes (people will say things like: 'It was at the time when X also happened'); and people's responses ('We decided to move to town X').
- 2 Ask these same questions of a lot of people. In the statistics section, we discuss samples. Here, you are compiling your own sample: be sure it is big enough to carry weight and representative enough to express all interested voices.
- 3 Ask the questions precisely, seek concrete detail and record the answers accurately. Look at the chapter on interviewing for hints: this is one kind of interviewing where closed questions can be useful for getting definite answers, although you need to seek more expressive, nuanced responses too.
- 4 Your answers can then be used to build your own database.

The value of structuring your questions

You want to know how the disappearing forest has affected people's lives. Asking "How has this affected you?" will provide the human, emotional answers: great for including in your story. But adding something more structured will give you database material, for example:

- 1 "Before they began cutting the trees, why did you visit the forest?"
 - Gathering wild plants and fruits
 - Hunting
 - Collecting honey
 - Collecting traditional medicines
 - Traditional worship
 - Collecting firewood
 - Collecting wood for building
 - Collecting wood for crafts
 - Other... (say what)"
- 2 "Which of these was your most important reason for going into the forest?"
- 3 "Which of these can you still do?"
- 4 "How has this affected your diet/ your daily life/your income/ the culture of your community?"

These answers will allow you to build up a land-use database on the role the forest played in the village economy, as well as assessing precisely which areas are now suffering. But, of course, you'll also want to ask: "How does losing access to the forest make you feel?"

Exercise #1**Creating your own database**

Draft five structured questions that will help you establish a mini-database on how a remote rural village has been affected by the influx of construction workers who are converting a nearby lake into a dam. You've heard that they have money to spend, and so can pay higher prices for fruit & fish, but also that there have been many more incidents of assault, rape and fights at the beer-hut.

Questions

COMMENTS

There are many ways to do this. But ideally your questions should focus on a before/after framework, focusing on economic and social life. So five 'starter' questions might be:

- 1 Do the dam workers come to the village:**
 - To buy food?
 - To buy drink?
 - To socialise?
 - Other?
- 2 (For village traders) What business are you in?**
- 3 Since the dam workers started visiting the village, has trade at your business**
 - Gone up?
 - Gone down?
 - Stayed the same?
- 4 Can you estimate roughly the value of this change?**
- 5 Since the dam workers started coming to the village, would you say**
 - There is friendliness between workers and villagers
 - There are tensions between workers and villagers
 - Can you give me an anecdote to illustrate your judgment about this?

...and so on

When dealing with accusations, or issues of responsibility – ‘This land developer from overseas was given our land by the government and now we are landless’ – it becomes even more important to have very precise statements from a large group of people. (After all, the land deeds register is in disarray). Interviewees will have to give as much detail as they possibly can (‘The land that stretches from the edge of this rock formation to the river in the South was given to my family by Governor X. Great-grandfather often talked of his grandfather meeting the Governor, a man with a big moustache’). An internet search on the history of colonial times in the country may yield a description or picture of the governor with the big moustache: an indication that there is truth to the claim. (Thank goodness for the internet, which reaches even where there are no papers!)

Interviews with this large group of now landless people will add up to a long list of similar stories, showing that at a certain point in time they were visited by people from a certain company, showing them papers to say that they were now evicted. Even if they were not given the papers, some will know the name of the company.

The rest is easier. If the company is an overseas land developer, it will have a recorded existence. Even if the national company registry is in as much disarray as the land deeds archive, international websites will yield details. Many Western company registries are accessible on line; the South African registry can be accessed through FAIR, and in white or yellow pages and also online. You can then phone the company and ask for comment – and also ask the country’s Land Affairs Ministry why they allowed the rightful owners of the land to be evicted. And there’s the story.

If it’s a local land-grabber, ask employees in the area, now in charge of fencing off the land and preparing it for development, to identify the ‘Big Boss’, if not by name then by physical appearance, language, accent, or car number-plate. Relentlessly bothering local (licensing) authorities and phoning the Land Affairs Ministry will eventually turn up a name.

Case study: Food scarcity in Zambia

No matter what issue attracts your interest, local people have a lot of knowledge in their heads to start off a hypothesis. Agricultural editor Chali Mulenga in Zambia’s southern province, for instance, already has a fair idea of the reasons why his province is no longer producing so much food, although at one time it used to be the country’s ‘breadbasket’. He sums up his hypotheses:

‘Development programmes not working, HIV/Aids unaddressed or badly addressed, cattle disease unaddressed, free food help from charities. All these elements have made farmers either unwilling or unable to continue farming.’

Without municipal or even national paper records, the statements of the people (Mulenga’s observation and own mini-survey) provided sufficient substance to create a hypothesis around the causes of the present food crisis.

Mulenga continues: ‘To do a story, I’d visit a number of previously agricultural areas in my province to observe and investigate, through interviews with former farmers, the correctness of the hypothesis. Then, in case the government hasn’t any documents to give me, I’d get the documentary records from development organisations and charitable donors. I’d then question government officials and the donor agencies. I would have enough to go on to produce the story.’

Mulenga’s only problem: lack of funds to travel to test his hypothesis in the various provincial areas he has in mind...

Basic research – skills and tools: following the paper trail

The phrase ‘paper trail’ is a metaphor, derived from the school racing game where a leader ran through countryside dropping bits of paper, and the following group tracked him as fast as they could by finding and following the bits of paper. If you think about the metaphor, you can see in your head what a ‘paper trail’ means in investigative journalism too.

Following a paper trail is the process of

- Identifying where the documents you need to back up your investigative hypothesis are
- Developing a strategy to access them
- And then using one document to lead you to the next relevant document.

For example:

You are trying to track down someone’s reputation – but all you have is a birth certificate and a CV

- You search more widely for records about the person. This is called profiling. Specific databases can help you to profile a person:
 - the company registry, where you can find out which companies, if any, someone is a director of, which then can lead you to this company’s shareholders and shareholders’ annual reports
 - the (land) deeds registry, where you can find out which houses, properties or lands someone owns
 - court records that can show whether someone has been involved in a court case
 - licensing authorities, where you can find what car they drive
 - tax authorities, where you can find out if a businessperson is registered for VAT
 - and if you want to see if the street addresses belong to posh villas or more modest abodes, there’s always Google Earth, which allows you to zoom in on an address from a satellite photograph.

- Then you follow the paper trail back to look for links between your findings and what you know of their work record.
- You jot down everything in these starting documents that might be relevant. This could be the fact that the person's CV says he was employed as a 'security officer' for a mining company in an area that was affected by civil war (involving conflict diamonds) in that period. Or it may be that a few years are not covered by any information at all. (This is where basic number skills come in. Add up the years worked at various jobs, add up the total years reflected in the employment record, and see if there are gaps.)
- You may find that the work or company records indicate the person left one position very suddenly. This would lead you to look for more documents about that person at that workplace or company. You might find that the human resources department recorded a complaint of theft or fraud against the person. You could follow that bit of paper to look for police/court/prison records about that person – and so on. In other words, you use one document to lead you to another, and provide reinforcing confirmation
- So you can distinguish between relevant and irrelevant documents, you will need to exercise empathy again. Put yourself in the subject's shoes and picture possible scenarios: what might he have done then? Why might he have done that? Would it make a difference if he chose option A or option B? This will help you avoid wild goose chases. If, straight from having been abroad (under a bit of a cloud) for five years, someone is appointed as a 'consultant' to the President, it doesn't make sense to look for papers about the person's history in the President's office. The appointment probably happened at high level and behind closed doors, based on, at most, a simple, short-term contract with few details. Instead it's likely to be more productive to look for information in the overseas country where the person purportedly stayed, or to try to track their cross-border movements during their absence.

Much of the paper trail can often be followed through public-record documents, though you will have to use your source-cultivation skills to access privately-held papers.

A useful description of a technique known as 'parallel backgrounding' can be found in the *Investigative Reporter's Handbook* (IRE). Parallel backgrounding is the method of comparing paper trails – for instance one on a company and one on a manager at the company – to find interesting facts. For example: the manager's records won't show you that a building that he managed for that company was closed down by the city council for illegal gambling, or that he headed a questionable tender bid. You find those facts when you compare the company's history with the manager's personal career.

Libraries and local newspapers

Many journalists think that using libraries and archives is simple. They are indexed alphabetically. You just look for the person's name. However, it often is not as simple as that. If you work with computerised records, entering a name will very often pull up relevant results – although often alongside a lot of irrelevant ones.

But in many African countries, public records are simply filed boxes of papers stacked in a dusty room. As well as negotiating with the gate-keeper who controls access to this room, there is another basic principle you need to grasp:



Before you search for a document, find out how the documents are indexed, and how to use the index!

This can save you hours.

Don't neglect news databases. Very often searching for public-record documents such as birth certificates or driving licences seems like the best way to start. But on-line news databases can also be surprisingly useful even for searches about a person. If someone is using their real name, a news search can turn up court cases they have been involved in, or even apparently trivial information like their attendance at a university function. Each of these fragments of information provides a piece of the jigsaw puzzle of their lives you are solving. An excellent news database that will churn out international reports on keyword search is <http://newslink.org>. You can also try www.topix.net.

Nowadays, many local newspapers are also available online. Don't ignore them and don't ignore your own local newspapers. They often contain massive amounts of profiling and paper trailing information, such as:

- information on local structures like banks, companies, government offices;
- information on individuals and their social networks, from reports on marriages, funerals and obituaries to family relationships;
- paid legal notices on wills, name changes, foreclosures, auctions, tenders, seized properties, unclaimed property, new construction projects;
- arrests and convictions of law breakers: you may find the name of your school bus driver in a drunken driving case, or the name of a school governing body member in a rape case.

Profiling character

When profiling a person, our newshound instinct will hope for the best, which is often the nastiest, story. But real facts are what you need.

How do we get a 'feel' for the human being?

- Start with what assets he owns. Very few hardworking honest African citizens have mansions in Monaco.
- Then examine his personal history. Did he forego personal enrichment opportunities to stay with an institution, like an inner city hospital, that needed him? Do his children attend state schools?
- Finally, talk to people who know or have worked with him. Do people who know him describe him as a good man? Faced

with such evidence, you may begin to doubt if you are on the right track with the nasty headline. (But beware of ‘inner circle superlatives’: many big crooks are surrounded by sycophants, who will gush relentlessly about his intellect, his love for the people and his many good deeds. That is **not** the kind of character reference you should immediately trust.)

Keep in mind that starting with an issue often leads you to a person or group of persons, and profiling databases work the other way too: a piece of land or a street address can lead you to the owner.

To sum up, you follow a paper-trail by:

- Web-searching, visiting archives, and persuading sources until you have assembled whatever public-record, open or easily accessed documents you can find about the person
- Mapping these, using the data-mapping techniques we describe in earlier chapters
- Looking for gaps, contradictions and inconsistencies
- Then thinking about what documents could fill the gaps or resolve the contradictions – and start looking for these.

Exercise #2

Follow a paper trail

Take some easily-defined and clear government or local government function – for example, providing school meals, or awarding hospital supply contracts – and map what paperwork needs to be done at each stage of the function.

- What forms have to be filled in? Who signs these? Where are they lodged?
- Are they available for public inspection? What’s the procedure for accessing them?
- What rules govern the process? What counts as ‘breaking the rules’?
- What’s the budget for this function?
- How has that budget been spent over the past couple of years?
- Who has got the contract? Is it the same company in successive years?
- If the supplying company has changed, why?
- If there was misconduct or unsatisfactory service, what was done? Was there a penalty?

You may find a story. Even if you don’t, you will have become familiar with the processes and rules that govern one aspect of local government in your area.

Basic research – skills and tools:

database management tools

If you have access to a good computer and software, there is an electronic project management tool that enables you to create a database of interviewees, knowledgeable contacts, informants and their area of expertise. It also contains a catalogue of questions to put to them, fact-sheets to help you record established facts as well as assumptions that still need to be proved, plus hyperlinks to relevant documents, facts, statistics, databanks, minutes and interviews/talks.

If you are dealing with highly sensitive material and are concerned that state security or other hostile agencies must not get hold of this information, you need either to use code names and keep your key to the codes in a safe place, or keep your electronic copies of the tool on flash drives which can be kept safe in bank safes and which do not need to be saved on installed hardware.

But if you don’t have such high-tech resources, you can establish a similar system using hard-copy documents. It is best to keep separate notebooks for each story and even for certain aspects of the story. Sensitive documents need to be kept in safe places. Some journalists store multiple copies with friends and in other safe places, but others are cautious about doing this. The more copies you keep, the more likely it is that one of them will fall into the wrong hands.

A filing system helps to keep track of facts and information and enables you to access them quickly when needed. Keep a summary of the filing system in a safe and separate place.

Chronology building

Keep an eye on the chronology in your investigation. This does not mean that you have to present your story in a chronological way, but that you ‘pin’ found facts on a timeline of the events you are researching. This will help you to build a clear picture of what came before and after what, and what happened simultaneously. An example from Brant Houston: X says he wasn’t in the country when a certain event happened, but, months later, you find minutes of a neighbourhood committee meeting that took place during that time, and which X attended. The inconsistency could easily escape you if you didn’t pin the statement from the interview and the meeting date from the minutes at the same point on the events timeline.

Computer assisted reporting (CAR)

Computers and the internet have dramatically increased the scope of information available to reporters and the ability to organise, retrieve and analyse it.

Basic research – skills and tools:

internet searching - Google and beyond

Search programmes have revolutionised finding data on the web. The best known is Google (www.google.com), but there are others, such as Yahoo (www.yahoo.com) and “meta-crawlers” which do the same search on four or five search engines at the same time.

The trick to efficient web searches is to choose your search keywords and phrases with enough precision to exclude the masses of results that are irrelevant to what you are interested in.

1 Set your preferences to return the maximum results

The Google page has a link marked ‘preferences’. This allows you to set some search preferences (to search only English language pages, for instance) but the most useful preference to change is the number of results shown for each search. The default setting is for 10 results which means you have to refresh the page each time to get the next ten. Set preferences for the maximum 100 results – which allows you to scan much bigger chunks of data to see if it is relevant.

Keywords provide a simple way of narrowing down your focus, but, often, keywords alone are not enough. Say you are looking for John Smith. Simply typing in John and Smith in the search bar is going to give you every document where both of those names appear: hundreds of thousands of documents. To avoid drowning, you will need to find relatively unique features that identify the John Smith you are looking for.

2 Use quotation marks

“John Smith” will return only those results where the words appear together. If you have a middle name you can add that, for example:

“John Sylvester Smith”

You can combine options using the OR command written in capitals, which Google uses to distinguish from the word ‘or’.

“John Sylvester Smith” OR “John S Smith” OR “JS Smith”

3 Add facts that you know or suspect

Say the John Smith you are interested in is alleged to be involved in drug smuggling and operates out of Zurich.

You would add to your search bar:

“John Smith” Zurich

or perhaps

“John Smith” Zurich drugs

Which would return only pages where all those words occur.

4 Country-specific searches

You may not be certain that Smith operates out of Zurich, but you are sure he is based in Switzerland. Using the “site:” command, Google allows you to search only pages with a specific country domain designation. The Swiss domain designation is “.ch”

So you might type in the search bar:

“John Smith” site:.ch

which would return all Swiss pages containing the name John Smith; or

“John Smith” drugs site:.ch

The domain designation for South Africa is .za, for Britain .uk

Not sure of the country designation? Google: “domain by country”

5 Organisation-specific searches

Many commercial websites end with .com; many NGO’s, developmental organisations’ sites and activist groups end with .org.

So if you are researching wind turbines, and want the companies, you might use

“wind turbines site: .com”.

If you want to find criticism about wind turbines, you might use

“wind turbines site:.org”.

If you want data on activist groups in South Africa, you type

“wind turbines site: .org.za”.

6 Use the net to find sources

Drug smuggler John Smith may never have appeared on the net in that context (as an accused drug smuggler) so the next best thing is to find an expert on the drug trade in Switzerland who might have heard of John Smith and be able to give you more information.

“drug smuggling in Switzerland”

or

“drug smuggling” site:.ch

should give you access to newspaper or academic articles giving the names of such experts.

You can then google their names to find their telephone numbers or email addresses and make contact.

7 Using automatic translations

Your articles from Switzerland may be in German or French. Google results offer an automatically translated version which can give you a good sense of what the article says (click on the note ‘translate this article’ under the result), but note that this is a very inexact process and you may have to struggle to make sense of the machine-generated translation!

8 Using Google cache

Web pages change or are shut down. You may get a result on Google and find the page has gone. Then click on the “cached” link on the specific result. Google saves a copy of the pages that it catalogues as it searches the web, and that is the cache version: the snapshot of the page as it was when Google’s computer looked at it. That copy is often still available long after the original page has disappeared from the net. This is very useful for tracking companies and individuals who have ‘disappeared’: they often still exist in caches.

9 Finding databases that are not covered by search engines

Many useful databases are not covered by Google. This includes many newspaper archives and municipal property databases and (in some countries, such as the USA) court archives. In looking for traces of John Smith, it might be useful to access Swiss newspaper sites and search their archives. Most archive searches work the same way that Google does. You can also type in the URL of the archive you are searching in the ‘domain’ box that appears when you click on Google Advanced. In that way you can get Google to search that specific archive.

10 Use the internet’s phonebook

Nearly every country has an extensive telephone database, usually under the term “white pages” (even many non-English-speaking countries list “white pages” of their phone directories in English).

So, for example, to try to look up John Smith’s Swiss number, you would Google

“white pages” site:.ch

and find the sites that offer Swiss telephone directories. Generally, directories require that you specify at least the town/city as well as the name.

11 Download long articles for later reading

If you have limited opportunity to go online, then save pages that look useful for background research so you can go through them carefully later.

12 Build up your own database in a structured searchable way

When you save documents from the internet, or save transcripts of interviews or notes, do so in a way which will allow you to find information again easily, or your virtual desktop will end up like many journalists’ actual desks: a vast, widely spread pile of assorted data where it is difficult to find anything at all, let alone quickly.

There is a great free internet tool, called ‘Google Desktop’, that searches and lists your computer files for you. Simply key in ‘John Smith’ and the tool will give you a list of all the files you have saved, even many years back under you can’t remember what name, with ‘John Smith’ in it.

You can also be more efficient by doing these four things:

- Date your documents and notes in the title, beginning with the year e.g. 20070327
- Change or add to the title of the document to include keywords that will be easily searchable and tell you something about what is in the document. So, to take our example, the academic study on drugs traffic in Switzerland which describes someone you have identified as John Smith might be ‘20070527 john smith swiss drugs backgrounder’
- Organise your research into folders. Create a John Smith file into which you can put all your gathered information.
- Build a chronology.

Taking these steps adds a bit of time in the beginning, but means documents are easily retrieved and you can roughly tell what’s in them without having to go through them all over again. In the long run it’s a time-saver.

A digital template for database management can be found at <http://www.luuksengers.nl/training/login/>. Though this is a

site from the Netherlands, and in Dutch, you'll find a useful example (in English) of a digital filing system if you click on "digital file template". A click on "Manual" will give you a good idea of how to use it.

Lastly:

Remember that for many of your readers, computer-researched information is a mysterious field they may not be able to explore for themselves. For this reason, bear in mind the following ethical points about stories based on CAR:

- be transparent about the data you find and use. Where it is possible, publish detailed references, or links to sites where original documents can be read
- verify your data very carefully, including checking the date of the information
- draw the correct conclusions from statistical and numerical data; your readers may not be able to do the calculations and have to trust your maths.

Basic research – skills and tools: data-mining

Data-mining is arguably the most objective process to help you arrive at a story choice. Think about it: which lead is more likely to put you on the right track: a complaint from one hospital patient about thieving nurses, or a database from the Health Ministry on how many disciplinary hearings and dismissals were the result of complaints about theft by state hospital workers over the last five years?

As with all information, we should always be mindful that even statistics can be manipulated and used to misinform, rather than inform (more on this below). But efficient 'mining' of databases has generated immensely important stories in the last ten years.

- The Danish CAR institute Dicar analysed payouts of state farm subsidies in the country over a certain period of time, and found inconsistencies and indications of favouritism that made headlines. Google search farm subsidy.org to see the results.
- *The Washington Post* collected freely available data on donations received by political parties and was able to headline the story "Super rich step into political vacuum" <http://www.washingtonpost.com/wp-dyn/articles/A38722-2004Oct16.html>.
- The Society of Environmental Journalists in the USA collected and compared pollution data in California and discovered that pollution in a certain area had become much much worse over a number of years. The SEJ has published a detailed report on what they did, how they did it and how they used a spreadsheet, Microsoft Excel, to calculate the results. Find the online tutorial on using spreadsheets by Russ Clemings on <http://www.sej.org/resource/tools.htm#online> and <http://www.sej.org/resource/tools.htm#online2>.

Of course, to have data to mine, you need a data-rich environment. In the USA, many institutions routinely file figures, statistics and graphs on what they do. For example, the flight safety authority will publish figures of complaints, alerts, near accidents and real accidents every year. You only need to combine and analyse these figures, using a spreadsheet to find, for example, how many accidents happened in 1998 as compared to 2008, and already you have a story: "US airspace as safe as never before", (or, depending on the findings, "US airspace more dangerous than ever before").

Census statistics, when mined, can give incredible results: "The following towns now almost exclusively inhabited by Hispanics", for instance, or "No African-Americans left in Mississippi" (OK, we made that one up!)

In Africa, we can only envy all the resources our US colleagues have at their disposal. But we too can sometimes use data-mining. There is more data available than many journalists realise. Journalists anywhere can sign up for regular email alerts from Stats SA: www.statssa.gov.za or look at www.sairr.org.za

International data can give results that are relevant for Africa

For example, development aid donors will publish reports on how they spent their money in any given year. By collecting these data from the donors that are most active in your country, putting them in a spreadsheet and analysing them, you can achieve results that can make headlines in Kinshasa – "Donors (to our country) spent most aid money training our civil servants", for instance. And it doesn't always have to be about money either. Collecting reports on French official visits to your country might give you a name of a well-known arms trade middleman, who always tags along.

Focus on social networks

Social networks can be members of a certain profession, members of a geographic community or prominent people in a certain political party. You can combine data on how much they earn, who they work with, who they have public meetings with, and end up with a social network picture that tells you something about their influence in society. Social network analyses have produced stories on terrorist networks, political party supporters, and the most influential or richest people in a certain geographic community. Find a free downloadable manual on <http://www.ire.org/sna/#links>.

Compile the necessary data yourself

You can, for instance, check all the tenders offered by your government over the past year, or three, or five years, and check which

companies received them. Were they the cheapest? The best? Or the ministers' friends?

Use the databases of journalists' and other organisations

In the USA and Europe, investigative journalists have established centres that produce databases for mining by journalists. Nicar in the USA, for example, collected data on Guantanamo Bay terrorism suspect detainees and compiled a database that can be accessed online <http://www.nicar.org/downloads/>

It will probably still take some time before archives in African countries are properly maintained, let alone become available online, but investigative journalists in countries where donors are active in the digital media and information field can lobby these donors for database projects for important archives.

Access to information

In many countries, especially in Africa, government information, or private sector information that impacts on citizens, is still kept away from the public, shrouded in Official Secrets Acts or simple unwillingness.

Governments and the private sector in the rest of the world are not always that much more willing to open up. It took seven years, from 2000 to 2007, before a number of European governments finally succumbed to freedom of information pressure on the issue of business subsidies. In order for the European public to find out where state subsidies for businesses were going, committed journalists in six countries had to work together, and support each other through court cases, to secure the release of the information.

The results were more than worth it. The list of main recipients of state subsidies turned out to be headed by captains of industry and members of royal families. They, rather than small or struggling businesses, had received millions of pounds' and Euros' worth of taxpayers money to subsidise their already highly profitable enterprises. Find the story on <http://www.guardian.co.uk/media/2007/jan/22/mondaymediasection.freedomofinformation>.

The fact that even in the West, where open records laws now do exist in almost all countries, it remains a struggle to use these laws effectively, shows that journalists will always have to work hard to get the information they need. A law only means that a door can be opened: you still have to find your way to the door and knock on it until it actually opens. And you have to know the relevant laws in detail in order to do that. In Chapter 8 we look at the general effectiveness of freedom of information (FoI) laws for journalists.

If you don't have an open records law, and are possibly still plagued by the existence of an Official Secrets Act, you probably struggle every day to get any public or private sector information at all. You may also be pestered by civil servants who will only give you documents in exchange for money, knowing that you won't otherwise be able to access them.

Paying sources for interviews or documents is not good journalistic practice. Most ethics rules and guidelines forbid it. The reason is clear: you can never be sure if the information was given in the public interest, since somebody had a financial reason for wanting to pass it on to you. If documents and interviews are worth hard money, people are encouraged to say or photocopy things for monetary gain – or even fake evidence for money. This is not how good journalism is supposed to work. We discuss these issues also in Chapter 8.

Yet how do we avoid paying for documents if there is no other way of getting the information that we need? The long, hard way is through struggle for access to information legislation and practice.

The SADC's Windhoek Declaration, The African Declaration on Human and People's Rights, and various African Union statements all advocate press freedom and the need for transparency in democracies. The majority of African governments have aligned themselves with these lofty ideals. The struggle for journalists – and in fact for all ordinary citizens who want to know what the powers that be in their countries are doing – is to exert pressure until these powers are forced to put, so to speak, their information where their mouths are.

Arguing for access to information

In 2005, the Zambian government rejected proposed freedom of information legislation, on the basis of two arguments

- that access to state information would compromise state security
- that 'there is press freedom in Zambia' already.

These two arguments are not difficult to counter. As MISA Zambia at the time protested: how do countries like the USA, who have access to information legislation, protect their state security? Aren't their security interests bigger than Zambia's? Any access to information law protects, for example, military secrets. The argument that **all** information needs to be secret because **some** information necessarily has to be, is therefore false. The second argument, that "there is press freedom", holds even less water. With a government and a public sector under no obligation to facilitate information to the public, what is press freedom? What are media free to publish?

Journalists and members of the public in Zambia, continue to engage the government and government officials and private sector representatives on this issue. Many press freedom organisations, media institutes and journalists are engaged in such struggles all over Africa. FAIR closely monitors their efforts and facilitates networking between access to information activists in the various countries. On behalf of its members, the FAIR help-desk also lobbies governments and companies directly to provide access to their records.

FAIR also encourages its members to join committees of access to information activists and to publicly request information, and document the responses to such requests. The FAIR website's Access to Information chapter reports on the progress made in the African struggle for access to information as well as on existing information laws and how journalists can make (better) use of them. A report card on the "Most Secretive Government" is also envisaged.

Using access to information laws

Using countries with free information access

Article 32 (1) of South Africa's Constitution states that everyone has the right of access to any information held by either public and private bodies that is required for the exercise or protection of any right. The Promotion of Access to Information Act (PAIA) gives legislative expression to this right, although private companies have been given an extended timeframe to come into conformity with the act.

The South African History Archive (SAHA), located at the University of the Witwatersrand, has built up impressive experience in using this law. Many journalists battling to gain access to government and private sector information have been helped by the archive. Studying the South African experience may be helpful to the struggle for access to information in other countries.

SAHA's Freedom of Information Programme is dedicated to using the South African Promotion of Access to Information Act to extend the boundaries of freedom of information and to build up an archive of materials released under the act for public use.

This is relevant to other African countries, too: South African government departments and SA businesses hold a lot of information that impacts on the southern African region, and even further north. For instance, the SA headquarters of Checkers has information on the supermarket group's activities in Zambia; the SA military holds information on SA military activities in other African countries, past and present.

By getting information relevant to your country from a South African archive, you can start a paper trail that you can follow up in your own country. It is therefore useful to remember SAHA when you follow a story with more than strictly local links. (The same applies to general investigations of foreign companies or institutions active in your country: their headquarters might be located in a country that has an access to information law, and you can use the FAIR network to ask a colleague from that country to access information on your behalf.) Dedicated organisations in the USA, the UK and elsewhere will also help if approached to use FoI legislation existing in their countries

How to work with SAHA

SAHA wants to develop its work within the region and is willing to work on joint projects with individuals and organisations seeking access to records that might be in the SA archives under its very broad mandate of documenting and researching "struggles for justice" (historical or contemporary).

SAHA will assist non-South African nationals in such research. There may be certain costs attached to specific investigations, but generally, SAHA does not charge unless there are costs involved in securing the documents (for instance, copying costs).

If the project is large-scale, SAHA may look at trying to raise specific project funds, jointly or independently. This will depend on the request made and on any discussions held about it.

SAHA believes it should be possible to secure documents from the SA Defence, Intelligence and Foreign Affairs departments as well as other ministries dealing with Trade and Industry.

To ask SAHA for help on line, go to www.saha.org.za.

- 1 Select INFORMATION REQUEST at the bottom left of the screen, or on the homepage and fill out FORM B: PAIA REQUEST.
- 2 Fill in the necessary fields, being sure to provide both your I.D. NUMBER and your CONTACT DETAILS.
- 3 Click the SUBMIT button at the bottom of the page.

Using access to information laws (cont.)

- 4 The Programme Manager will respond to your request and if necessary have further discussion about it.
- 5 SAHA will then draft and submit a request on your behalf to the relevant body.

Results in South Africa

After four years of experience using PAIA, the programme has built up a comprehensive archive of released materials and has put considerable effort into spreading awareness of the importance of access to information in a transparent and accountable democracy. There are now new initiatives to spread the lessons of this experience and knowledge about it, so that ordinary citizens are empowered to access information that has a direct impact on their daily lives.

If your country does have FoI laws, the following general principles should guide your attempts to access information by these routes:

- Always check first whether the information is already 'out there'. Limited-circulation published papers sometimes contain summaries and even extracts from supposedly secret documents. Find out what has been published in semi-official or specialist contexts on the subject, and try to find a mole who will let you see the relevant paper.
- Use FoI provisions as a last resort. If you can demonstrate that you have genuinely tried every other channel, this strengthens your case for demanding the paper.
- Plan ahead: FoI procedures can be slow and you're very unlikely to get a paper you need tomorrow!
- Identify and approach the right information-holder.
- Make precise requests for named (or numbered) documents. Asking for "everything you've got on..." will not get results.
- Document your requests and the responses you got very carefully. You may need these records to prove that the authorities are deliberately flouting FoI laws and have something to hide.

Basic number skills

Much investigative journalism is qualitative: it looks at why and how things go wrong and who might be responsible. But almost every investigative story has – or needs – an underpinning of numbers: how big was the deficit; what are the statistics on illegal fishing; how many patients are turned away from clinics and how do we know?

So you need to know how to tell a big number from a small one, how to make sense of numbers, and a few simple things like how to calculate a percentage. No one becomes a journalist because they really love numbers, on the contrary. But they are not difficult and are essential.

Many people who believe they have no skill with numbers actually use numbers in quite a sophisticated way every day: budgeting for living expenses; working out whether a season train ticket represents good value; negotiating for a pay increase, for example. The way many schools teach number skills has contributed to the fear of ‘mathematics’ in perfectly numerate people; they have learned to disconnect the practical applications of numbers from the apparently abstract science of maths. The good news about number skills for journalists is that they focus mainly on application, and have a strong qualitative slant (for example, in understanding who collects statistics, and how and why). But you need to start by grasping the basics of numbers.

Exercise #3

Numbers and statistics

Look at the story ideas below. Decide for each what kind of numerical or statistical information is needed to strengthen the story.

- 1 A government minister has outraged nurses by making a speech in which she alleges: “Too many nurses are lazy. They spend their days drinking tea, while patients wait outside the clinics.”
- 2 Local shopkeepers are complaining that a major construction site in the city centre is polluting the air and threatening the health of their workers and customers.
- 3 A prominent church minister has started a national campaign against what he describes as “the highest levels of teenage immorality in our history”.
- 4 Your national broadcaster has issued a press statement claiming that a recent TV series on family life “got the whole nation engaged with the issues”.
- 5 The maize farmers’ national organisation announces it will be raising prices because of “unprecedentedly adverse weather conditions this year”.

Comments

1 This is a story about a value judgement and attitudes, so numbers may not change minds. But if you can get a clinic nurse's official job description, and the help of experts who can put times on each task, you can create a 'typical day' timetable. Then, via observation and interviews you can find out:

- Are nurses faced with too many tasks to fit into their working day?
- What tasks occupy most time? Do the nurses use short-cuts? How?
- How does their job description relate to the average numbers of patients a clinic sees? (Go to your local clinic and count the people in the queue. Get the agreement of a patient to accompany her through the clinic processes. Time how long this takes and multiply by the number of people in that day's queue.)

None of this is rocket science – or even difficult arithmetic – but the figures you get will anchor your story in reality.

2 You need to get an air sample analysed (see if your local college can help if you are not close to high-tech facilities). Find out what pollutants are in the air, then ask a medical expert whether these are dangerous and what levels of exposure will damage health. Match the levels against any air purity regulations your country has. This is a story where figures can prove that the shopkeepers do have a justified complaint.

3 What does the minister mean by immorality? You can't look at this story in number terms until you can get a definition. But, once you find out (for example, he tells you: "teenage criminals in prisons") you can look back at statistics on prisoners under the age of 18. You may find that this problem started a long time ago and that the figures haven't changed much over time, or that similar 'peaks' in the figures seem to occur fairly regularly – or even that the figures are lower now than they used to be! This is a 'trend' story: somebody creates a stir about something, alleging that it is worse than it has ever been. The journalist's job is to interrogate the assumption, because very often all that has happened is that for some reason the problem has become more noticeable, rather than actually bigger. You need figures to unpick the assumption. But that isn't enough, because the context – why is the problem more noticeable (or noticed) now? – may be where your real story lies.

4 Press releases are not designed to be placed under statistical scrutiny – or at least, that's what their authors hope! (See the Joe Hanlon case study at the end of this chapter.) And this is another 'trend' story, although in this case inspired by self-publicity. But if you want to test just how influential the public broadcaster really is, select a sample group of typical viewers, and ask them some questions about whether they watched the series in question and whether it did get them into sustained discussions on the issues. Present your mini-poll results alongside your story.

5 Weather statistics have probably been kept in most countries for longer than almost any other types of figures. In Africa, they were among the first statistics the colonial authorities recorded, and can be traced back even further by community oral history traditions on floods and droughts. You can't just take what the grain growers announce at face value, because they obviously have an interest in justifying higher prices. So you must find out if the weather conditions really are 'unprecedented'. If they are, of course, you may also have a bigger 'global warming' story about your country's agriculture.

Look at the story below. It was written by a freelance reporter on a South African paper in the KwaZulu-Natal region, who read the publicity on a new credit card scheme aimed at lower-income families, and wondered exactly how the figures added up. (We have edited the story for context and length.)

Getting rich by helping the poor

GETTING RICH BY HELPING THE POOR

Banks make more than a billion rand a year lending money from the poorest-paid people in South Africa.

By Tom Dennen

The Consultative Group to Assist the Poor (CGAP) arrived on the scene in the wake of the semi-demise of the so-called micro lenders – to help the poor with a South African banking product called Mzansi, launched in October 2004. Mzansi means 'south', which is where some more South African money is apparently heading.

Remember the micro lenders? Those bad guys who sprang up everywhere, operated from sometimes shoddy offices, charged enormous interest rates and have now been mostly bought or driven out of the temple by the commercial banks?

Absa, First National Bank, Meeg Bank, Nedbank and Standard Bank (as well as Postbank, which administers the product) now offer the 'Mzansi' account which opens a potential customer base in South Africa of an estimated 13 million people earning under R5 000 a month – the erstwhile micro lenders' old clients. This group of 'poor' people was previously cut off from the banking mainstream because banks believed it was unprofitable to service them (and publicly said so).

They were wrong. And it took them just one look at the old micro lenders to see the light.

According to *Business Day*, (30 August 2006), finance minister Trevor Manuel confirmed that the number of Mzansi

Getting rich by helping the poor (cont.)

accounts opened since launch had grown to 3.3 million in just two years. “No management fees are charged on Mzansi accounts”, the website and brochures claim.

Mzansi Account (from official website):

Amount	Credit Interest (% pa, un compounded – a one-off.)
R1 - R499	0.25%
R500 - R999	0.75%
R1,000 - R1,999	1.00%
R2,000 - R4,999	1.25%
R5,000 - R15,000	1.75%

EXAMPLE: Mzansi Money transfer costs at the Standard Bank:

Amount	Cash to cash	Account to cash
R0-R100	R13	R13
R100.01 - R500	R26	R21
R500.01 - R1,000	R30	R25
R1,000.01 - R2,000	R40	R35
R2,000.01 - R3,000	R60	R50
R3,000.01 - R4,000	R80	R60
R4,000.01 - R5,000	R100	R80

A sixty cent charge is levied for deposits made after the first free deposit. This means you get to lend Mzansi your money once a month (make a deposit) for free. But only once. After that it costs you sixty cents every time you lend them (deposit) more of your money within the monthly period. But there are ‘no administrative charges’.

At the time of writing, in January 2007, the average consumer of this new banking product was female, approximately 90% were black and each held, on average, R300 in her account. That’s an average of R300 times 3.3 million accounts – 300 times 3.3 million is – *voilà*: R990 million – close to a billion rand – from the poorest-paid people in South Africa! Just to round things down, 10% interest on a billion is a R100 million that a little over three million people are generating. And at the princely sum of 0.25% interest per annum, (75 cents on your average R300 balance) that money only costs the banks a little over R2 million – yielding a R98 million gross profit.

Well, not really: the first withdrawal for each customer – at the minimum withdrawal cost of R4 from an automatic cash machine (ATM) – wipes out not only the (uncompounded) 75 cent interest earned for the whole year but immediately adds more serious money to the Mzansi coffers:

$R3.25 \times 3.3 \text{ million} = R10,725,000$ for the first month when the 0.25% interest is wiped out. In English: add nearly another R11 million profit to the R98 million so far.

After the first withdrawal, there’s no more interest, so if our 3.3 million people each make just one withdrawal a month for the next eleven months ($R4 \times 3.3 \times 11$) you can add another R140 million profit. R990 million at 11.5% is a cool R113,850,000 profit, at a cost of only 0.25% or R2,475,000.

Getting rich by helping the poor (cont)

Gross Profit:

R 113,000,000 on lending

R 11,000,000 on one withdrawal

R140,000,000 on one withdrawal a month x eleven months.

TOTAL: R 260,000,000

This of course assumes that each of the 3.3 million subscribers makes only 12 withdrawals a year and sits on the average R300 deposit, never asks for a statement or makes any other transaction such as a transfer. These figures are basic minimum, but just about guaranteed. And we haven't taken into account the cash to cash and account to cash transfer costs to customers yet.

But say we do: if each of the 3.3 million customers uses the account as it is intended to be used and makes one minimum transfer from or to the Standard Bank once a month – money home to family, perhaps - 3.3 million customers x 12 months x R13 transfer fee = R514 million. Add that to what we've got – R260 million – and here's over three quarters of a billion rand gross coming in every year. Cheap money. Shows thought and initiative. Buy shares now!

I wonder what the Mzansi CEO makes a year?

Things change, I know – guys used to wear tattoos, girls wore earrings and banks paid you (interest) to borrow your money. But there you have it: Mzansi borrows money from 'the poor' at a miniscule interest rate and lends it out at around 11.5% to – the rich.

The bottom line is that now even poor people can make money for banks and borrowers – and according to the CGAP, these new micro finance institutions (MFIs) are out-performing the commercial banking sector's Return on Assets Managed (RoAMs) by more than 50%. Globally. The whole world has discovered the poor and is taking 50% more from them in RoAMs than commercial banks are from the rich!

Here I would normally cry "Foul!" or "For shame!" but there is a big advantage here for the poor which is, of course, access to their own money. Nowadays the poor can transfer money to relatives without worrying about it being stolen in the mail, pay retail accounts with what is essentially a debit card and generally carry on as if their money actually counted and real interest was being paid for it, as 'normal' banking allegedly does.

I'd also like to get rich helping the poor but I've read somewhere that in Christian and Muslim societies usury is a no-no.

Interview with Tom Dennen



What questions were you trying to answer in the story?

The story is an examination of both a so-called 'credit' card system and the nature of financial communication in South Africa.



How did the story get started?

The story came about when I had to get banking details to one of the publications I write for – I had been running on a joint account. One of the cards on offer was the Mzansi. In terms of benefits, it seemed streets ahead of the competition until I learned about the no interest part – they charge very little, relatively speaking, but don't really pay anything.



What research did you do?

I looked at the Mzansi brochures from the Post Office, their website, the Post Office site and other bank sites offering cards. We also wrote to banks to try and confirm the figures and their profits. We received the following response from one:

"...We thank you for your query to which we kindly respond as follows. Mzansi is an entry-level first order bank account that has been priced on the margin and, after including the costs associated with providing the service, does not contribute to [this bank's] profitability as yet. Gross revenue is but one side of the picture. The reality is that the average monthly cost to the Mzansi customers based on their behaviour is less than R10 and customers only pay for the services they utilise. We also offer free cellphone banking options adding to the customer's convenience for non-cash transactions. The credit interest rate is a factor of the market and is higher than for traditional transactional accounts. Savers will experience capital preservation and appreciation meaning that R100 deposited will be worth more than R100 after a year. The banks offer significantly more savings and investment options for customers to earn more investment revenue. Mzansi remains an entry-level banking account. For any planned article, please quote..."

We didn't buy that spin for one nanosecond and published the article in the *Natal Weekend Witness* on the opinion page (p.20).



What difficulties did you encounter, and how did you try to solve them?

There were no real problems, just a 'routine investigation' as the cops say.

Getting rich by helping the poor (cont.)



How long did the investigation take you, and what were the most time-consuming and/or expensive aspects?

The story took about three months and most of the difficulty was in getting the figures right.



What lessons have you learned from doing this investigation and what advice would you give to other journalists attempting similar investigations?

Plod. My father-in-law taught me one of the best lessons of my life – how to plod. We were building at the time and he asked me: “See that pile of bricks over there? They have to be moved over to the other side of the property. Take two bricks in each hand and carry them over. Don’t think of anything much and pretty soon you’ll have those bricks on the other side of the property. It’s called plodding!”

Dennen’s story took a piece of bank advertising, and a concept – cheaper banking for the poor – that most people would instinctively support, and with some simple but painstaking maths, exposed that while it might have some benefits, it both did not offer full banking benefits to its customers, and helped the already rich to get richer at their expense.

It shows why every investigative journalist needs at a minimum some basic skill with numbers. Ideally, you need more than basic skills, but a good alternative is to know how to access these more advanced skills from others, either via a web resource, or by having a good relationship with a numbers expert in your locality.

This part of the chapter looks only at basic number skills. The resource list at the end of this chapter also recommends additional resources.

Basic research – skills and tools:

recognising and expressing numbers

This seems easy enough. Most of us know the whole numbers 1, 2, 3, 4 etc. and what quantities they stand for. You also need to understand what a minus (negative) figure is. Think of a thermometer where ‘0’ is freezing point. Minus 3 (-3) is three degrees **below or less than** freezing point. You need to be able to recognise fractions (1/2, ¼, 1/3 etc.), percentages (number out of every hundred: 50%, 25% and 33% for example) and decimals (fractions of numbers expressed in tenths and hundredths: for example 1 1/4 is equivalent to 1.25).

When writing about numbers, it’s also useful to be able to translate these for readers into natural frequencies: in other words, rather than telling your readers that “25% of the population” does something, it’s far more vivid and clear to tell them that “one person in four” does it.

Basic research – skills and tools:

calculating a percentage

You also need to be able to calculate a percentage.

A percentage is a number out of 100. If a factory employs 200 workers and takes on another 50, that is an increase of 25%. Percentages allow you to compare change over time (trends). It’s a simple formula to work out percentage change:

$$\frac{(\text{New figure} - \text{old figure})}{\text{Old figure}} \times 100 =$$

Apply the figures from our example:

$$\frac{(250 [200 \text{ old workers} + 50 \text{ new workers}] - 200 [\text{old workers}])}{200 [\text{old workers}]} \times 100 = \mathbf{25\%}$$

Try this again using these figures:

The population of Port Elizabeth was 834 000 and has increased to 989 000

$$\frac{989\,000 - 834\,000}{834\,000} \times 100 = 18.6\%$$

What if the population has decreased? The formula is the same, but the result will be different, because the base is different. The population was 989 000 and has decreased to 834 000

$$\frac{834\,000 - 989\,000}{989\,000} \times 100 = -15.7\%$$

Keep this example, but all you need to remember is the formula: new minus old over old times 100.

Basic research – skills and tools: **rates and averages**

If you are going to investigate in fields that use numerical and statistical information, you also need to be able to understand the concepts of:

- Rate
- Average

1 Rate

Rates allow you to compare like with like, apples with apples. The most well known are the inflation rate and the per capita rate.

The 'inflation rate' means how fast costs are rising **over a given period of time**. When government reports that 'the inflation rate has fallen this quarter' it is not saying that costs have fallen, or even that they have stopped rising. It is announcing that costs did not rise quite so fast over the last three months as over some other three-month period: perhaps the previous three months, or perhaps the same three months last year. Per capita takes population differences into account. For example, in 2002 the deaths from tuberculosis in South Africa was 53 per 100 000 people. By comparison, it was 117 per 100 000 in Somalia, the worst rate in the world, while in Swaziland it was 94.

2 Averages

These are ways of reducing a set of figures to one figure that can be used as typical.

The Blue Football team has 11 players, some of whom earn more than others. Their top striker gets R21 000 a week, the goalie R10 000, four players earn R2 000 and the remaining five R1 000.

- **The mean.** This is what most people think of as 'the average': we add up the 11 salaries and divide by 11 to get an 'average' salary – in this case R4 000. It's obvious what the problem with the mean is: it masks the real spread of salaries and isn't particularly close to what anyone actually earns. The mean is useful for certain mathematical operations; it is often less useful for journalists whose stories come from what is going on inside the aggregated figures.
- **The median** (mid-point). This would be calculated by listing all the figures in order, and selecting the middle one:
 - R21 000
 - R10 000
 - R 2 000
 - R 2 000
 - R 2 000
 - R 2 000**
 - R1 000
 - R1 000
 - R1 000
 - R1 000
 - R1 000

On a small sample like this, the median of R2 000 is fairly accurate as representing most people's salaries.

- **The mode.** The mode is the most frequently occurring value in a set of figures. In this case, again, this is R1 000 and is reasonably useful as a representative figure. It is a real salary that also carries useful information: that most people in the group actually earn this amount. What it still doesn't convey is that one salary (the stiker's?) is very much higher. And it would be helpful to know the sample size and just how big a section the 'most frequently occurring' represents. Look at this example:

A factory employs 500 people: 350 very low-paid labourers; 100 clerical and administrative staff and 50 specialists and senior managers up to the highest-paid boss. Each of the top 50 earns a different salary – but so does each labourer, because they are paid according to complicated piece-rates. But all the clerical and administrative staff earn the same wage, because it happens that this year they are all on the same grade. In a wages survey, they'd be the mode, because their salary figure occurs most frequently – despite the fact that the very low wages of the labourers actually dominate the pay landscape of that factory.

3 Dispersion

The range of our footballers' salaries was between R1 000 and R21 000. The range of salaries in our imaginary factory was, say, between R80 000 and R800. The second set of figures is far more widely spread-out (dispersed) than the first. And sometimes understanding the significance of a figure depends on knowing how widely the set it comes from is dispersed, or understanding the 'normal' dispersion of such sets.

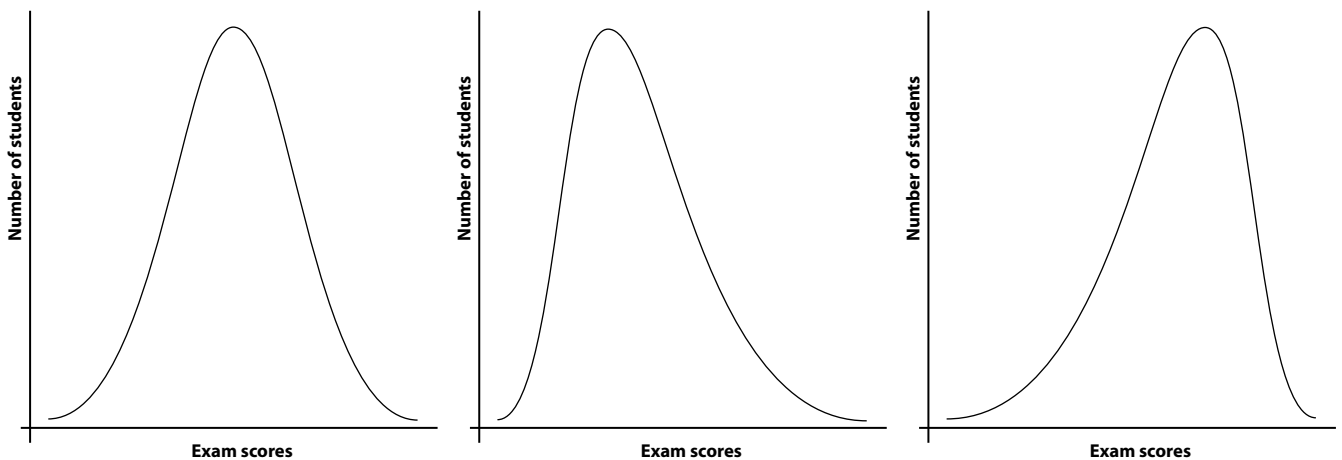
The simplest way of expressing dispersion is the one we've just used: describing the range ('from R80 000 to R800': a range of R72,000). Statisticians, however, use a calculating tool known as standard deviation to describe dispersion. They establish the mean of the set of numbers, and then calculate the average distance of all the numbers from this mean.

There is one regularly observed shape to the **pattern** of dispersion in sets of figures, whatever their range. It is referred to as the 'bell curve' because of its shape when the figures are plotted on a graph (a frequency distribution graph – below left).

As you can see, there are fewer results at the lowest value, they curve upwards towards the highest concentration somewhere around the middle values, then tail off again towards the highest values. This pattern happens so frequently that mathematicians describe it as a 'normal' distribution. When they don't see it, they ask questions.

So if, for example, a set of school-leaving exam results looked like the graph (below centre), they – and we as journalists – would be asking why such a high rate of failures occurred – because the bell curve tells us that under normal circumstances, the 'bulge' should be much closer to the middle (below left). Was the exam too hard? Was marking too harsh? Was teaching bad? Were teachers inadequately prepared or resourced?

Likewise, if the bell curve looked like the graph to the right below, the relevant questions would be the opposite: was the exam too easy; was marking too slack?



Both these sets of questions rest on two basic assumptions. The first is the general assumption that statistical science is reliable: that we can reasonably expect to see a normal distribution bell curve when we look at the pattern of school-leaving exam results – or any other set of data. The second assumption is far more education-specific and far less often held up to the light: that the purpose of examinations is to filter people so that some pass and some fail. If it were not, why would we worry when 'too many' do pass or fail?

These two assumptions are closely plaited together: 'what' is twisted up with 'how' and 'why'. This is because the science of statistics itself is very hard to separate out from the collection, presentation and analysis of statistics – all essentially human activities. The next section will try and untangle the different aspects.

Exercise #4**Check your number skills**

This short quiz is used by Derek Luyt, a lecturer at Rhodes University, to illustrate why journalists need to think about figures rather than simply quoting them unquestioningly. Take 20 minutes, read the information supplied for each question and try to answer.

1 Murder rates

The murder rate in South Africa in 2007 was 40 per 100 000. In 2007, Tsoko had a population of 10 000 and 5 people were murdered in the town. In 2007, East London had a population of 1 000 000 and 300 people were murdered in the city.

- Was the murder rate higher in Tsoko or East London?
- Was the murder rate in Tsoko higher or lower than the national rate? By how much?
- Was the murder rate in East London higher or lower than the national rate? By how much?

2 Electoral turnout

There are 42 000 people living in Dolo. 43% are under the voting age of 18, and 16 000 of these registered to vote in the last local elections. On election day, 12 000 people voted.

- How many people in Dolo were of voting age?
- What percentage of the people of voting age registered as voters?
- What was the percentage turnout on election day?
- Of those of voting age, what percentage voted?

Exercise #4 (cont.)**Check your number skills****3 Land Redistribution in South Africa, 1995 – 2004**

The table below was received in a press release from the national Department of Land Affairs:

Province	Hectares
Eastern Cape	36 200
Free State	37 320
Gauteng	8 115
Kwazulu-Natal	27 957
Mpumalanga	19 786
Northern Cape	43 251
Limpopo	23 591
North West Province	21 430
Western Cape	19 180
Total	262 520

- Which province redistributed the most land?
- Which province redistributed the least land?
- How much more land was redistributed in the Northern Cape than Gauteng?
- How much land was redistributed nationally on average in each of the ten years between 1995 and 2004?

4 'Average' salaries

Five journalists are discussing their monthly salaries. Jabu and Thembi each earn R1,200 per month, Bongani earns R1 600, Joyce earns R1 200 and Siphon earns R6 800.

- What is their average monthly income?
- How many of the journalists earn more than the average monthly income?

Exercise #4 (cont.)**Check your number skills****HERE IS DEREK LUYT'S COMMENTARY:****The main aims of Question 1 are:**

- First to discuss why crime stats are normally given as a ratio of 100 000, and not as a percentage – partly because it is difficult to envisage 0,03 percent of a murdered person;
- Second, to introduce the idea of ratios, which is expanded on in following questions;
- Thirdly to introduce how to work with decimal system (all the questions can be very easily answered without a calculator - you don't have to answer the second and third questions as percentages, higher/lower by 10/100 000 is fine).

Answers: Tsoko; higher – by 25%; lower – by 25%

The main aims of Question 2 are:

- First to go into greater detail on percentages and
- Second to consider which angle to choose/investigate, since the numbers don't speak for themselves.

Answers: Voter turnout was 75% (true) but the number of people eligible to vote who actually voted was 50,1% (true).

**Which is the stronger angle;**

a high turnout among those who actually registered, or the far lower registration rate among people eligible to vote?

The main aim of Question 3 is:

To check your ability to work out a simple average, but more importantly to stress the importance of checking your sources. The average amount of land redistributed each year was not, as you might total up, 26 259 hectares, since the number of hectares in column two **does not actually add up** to 262 520, but to 236 830 (hence average redistribution was 23 683 hectares/year). Lesson: check your facts and sources. This exercise is based on a real experience I had with the Department of Land Affairs.

The main aim of Question 4 is:

To consider the different measures of central tendency and distribution ('average'): mean, median, mode. Most people take the average to be the mean (add up all the salaries then divide by the number of people). Problem is, there are very rarely "average" people, nor does the mean always tell you anything meaningful about social conditions. In this example of income, the mean cannot tell us anything about income distribution. Most people in this sample earn **less** than the 'average' (mean) salary of R2 400.

How did you score? Did you spot the number traps in the questions?

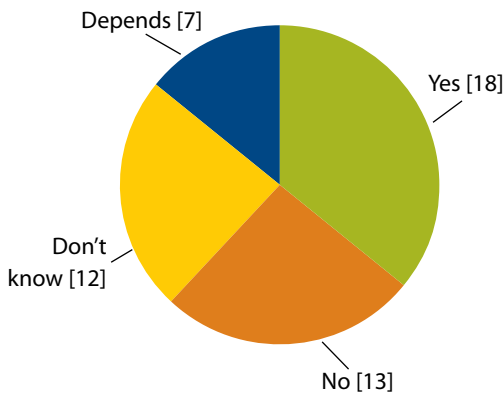
Basic research – skills and tools:**discrete and continuous variables**

There are some additional ideas about numbers that are slightly more complex. The first is that certain things we count – such as people, or anything else from potatoes to jet fighters that exists as a single separate and countable unit – are called discrete variables. That is, they move from one value – 'one potato' – directly to the next value – 'two potatoes'. That's why journalists regularly make a joke of the statistic that the average Western family has '2.4 children'. What would '0.4' of a child look like? Rather than explaining these unimaginable parts of units to readers, it's far more useful to round up or round down – in other words, to express the number as the nearest whole number. If the decimal is below 0.5, round down (so the average family has 2 children); if it is above 0.5, round up (if the figure were 2.6 children, you'd express it as 3).

However, when we are dealing with things that can be precisely weighed or measured rather than counted – for example, the weight of a loaf of bread, the distance between two cities or the volume of water in a dam – even our most accurate measurement is already only approximate. Water doesn't naturally exist in neat litre units; we're already rounding up or down (approximating) when we say a dam can hold 2 000 litres of water. Be far more careful in further rounding off the figures relating to these kinds of things, which we call continuous variables. Be sure that how you handle such figures reflects the degree of precision your story needs to make sense.

Sometimes, expressing numbers as words or realities readers can visualise (“one person in four”) helps readers. But sometimes assigning numbers to words can also help. If you’ve done a snap vox pop poll, it is quite heavy for readers to grasp in the middle of a story that “Of 48 people surveyed on the streets of Lusaka, 12 said they had no opinion as to whether SADC should discipline undemocratic member states, 18 believed the SADC Secretariat should take action, 13 believed it should not and seven said it depended on the circumstances.”

A far clearer way of expressing this would be with a chart or simple ranking order, like this:



We asked 48 members of the public, “Should the SADC Secretariat take action against undemocratic states?” Here’s what they said:

- 18: yes
- 13: no
- 12: don’t know/ no opinion/ prefer not to answer
- 7: depends on circumstances.”

Basic research – skills and tools: **creating questions from numbers**

It’s important that you ask precise questions about numbers – which means that you have to research and calculate in advance. These are the ground-rules for asking about numbers:

- ✓ **Wherever possible, quote the precise figures and their source in your question.**
 “Mr Minister, you say that grain stocks are adequate. But the Agricultural Union say that we have only 7 million tons of maize and that this is two million tons less than the country needs. What’s your comment on this?”
- ✓ **Use the ‘pinning-down’ technique: closed questions that demand precise answers.**
 “Bigger than what?”
 “How many babies died at the hospital?”
 “Is that an accurate figure?”
- ✓ **Ask questions in stages, not in groups, or your interviewee will only answer some parts.**
 “How much of the missing funds has your audit accounted for?”
 “Thank you. Please tell me where they traced the funds to?”
 “Were these payments authorised?”
 “And the money that has not been traced: what plans do you have for tracking that down?”
 “So when will this second investigation report?”

Exercise #5

Writing up numbers

What comments would you make on the following account of a survey? How could the story be improved?

Line 1	JOHANNESBURG – Yeoville remains one of the most crime-ridden areas in Johannesburg, according to a survey by consultant Martin Wessels for the Johannesburg Development Agency.
5	Crime had fallen from its 2003 high but remained problematic, the survey found. Public confidence was similar to the past two years but respondents were more optimistic.
10	Most worried about soaring assaults and burglaries, with 86% saying they did not feel safe outdoors in Yeoville compared with 55% in 2005. More than 80% said policing could not root out crime and 70% felt crime had worsened. Business confidence had changed little.
15	Compared with the rest of the city, 17% of the respondents said crime levels were the same, 32% saw improvement and 51% reported an increase. More than 70% said Yeoville had a dirty, unsafe image.
18	Only 6.7% felt there was not enough entertainment. A total of 56% expressed a need for larger retailers, 78% said street and pavement lighting was inadequate and 70% said that littering was a big problem.

The first thing to say about this story is that, although short, it is really hard to read. So many figures are crammed into such a small space that the reader is likely to get confused. But there are technical problems in the reporting too. We don't know how many people, and whether living inside or outside the suburb in question, were interviewed. So we have no context for their answers. We also don't know what questions were asked. Were people asked whether crime had fallen (Line 4) or is this information from another source? It seems to be contradicted by findings in lines 8 and 11. What does "public confidence was the same..." (Line 6) mean? Which answers does this relate to? Did the earlier survey question the same population and ask the same questions? If not, we can't compare them.

Is the 'business confidence' finding (Line 11) linked to the answers on crime with which it is grouped, or was another question asked about this? In Line 14, does the fact that "only 6,7% felt there was not enough entertainment" mean that the other 92,3% felt there was enough – or too much? And what do these answers mean? There's no attempt to either contextualise or analyse them, and the trends (difference from previous surveys) are not consistently noted.

Basic research – skills and tools: **writing figures up**

We've seen that figures (especially if they go to several places of decimals) may be very hard for readers to grasp. And numbers often 'slow down' a news lead or headline so much that it loses all impact. But inaccurately conveying numbers is actually lying to readers. Look at the tables below for a broadly accurate way of expressing figures in words. Be especially careful of the distinction between 'most' (which means a majority: more in this category than outside it) and 'many' (which simply means a significant number, but not a majority). 'More' is completely meaningless unless you say more than what.

And be extremely careful with 'at least' and 'at most'. The first means 'no less/fewer than'. So if you say 'at least four people were killed in a car accident' you are saying that you **know for certain** four were killed, and you are implying that there may be more. If you say 'at most' you are guaranteeing to readers that this is the highest possible figure. Don't use such expressions where there is uncertainty about answers.

Expressing numbers

Percentage	Summarised or re-expressed	In one or two words
100%	All	All
99%	Practically all	Most
95%	Almost all	Most
90%	Nearly all/Nine out of ten	Most
80%	The greater part or number/Eight out of ten	Most
70%	The greater part or number/Seven out of ten	Most
60%	More than half/Six out of ten	Most
55%	Just over half	Most
50%	Half/Five out of ten	Half
45%	Nearly half	Many
40%	A large part or number/Four out of ten	Many/A significant minority
35%	Quite a large part or number/Just over a third	A significant minority
30%	Roughly a third/Roughly one in three (remember 'a third' is precisely 33.3333 recurring)	A minority (significance here will depend on context)
25%	A quarter/One in four	A minority
20%	A fifth/One in five	A small minority
15%	A small part or number	Few/a few
10%	A tenth/One in ten	Not much/many/Few/a few
5%	A twentieth/One in twenty/A very small part or number	Few/very few
1%	A hundredth/One in a hundred	Very few/A tiny minority
0%	None	None
0.something%	A minute number or part/Fewer than one in a hundred	Almost none/none

Expressions and metaphors of number (get these right too!):

Metaphor or expression	Meaning in real numbers
Legion (comes from the numbers in ancient Roman army divisions)	4 000 – 5 000
Myriad (comes from the numbers in ancient Persian army divisions)	10 000 or a huge host. 'Myriads' would have to be at least 20 000
A horde/hordes	Huge numbers – has a negative and sometimes racist connotation
A few	Usually under 5
Several	6 – 9
A dozen/dozens	12/approximate multiples of 12 below about 50
A baker's dozen	13 exactly
A score/scores	20/approximate multiples of 20 below 100
Hundreds	More than 100, fewer than 1 000
A handful	Vague: usually below 20
A bucketful	Vague: usually above 20 and below 100. Used only of things, not people.

Statistics

Damn journalists and statistics

DAMN JOURNALISTS AND STATISTICS

by Peter Wilby

The Guardian, Monday November 5 2007

(This extract shows how important it is for journalists to get their figures right. Find the full story on the *Guardian* website: <http://www.guardian.co.uk/media/2007/nov/05/mondaymediasection.pressandpublishing>)

Over the past week, we have learnt that ministers have underestimated the number of migrant workers entering Britain during the past decade by 300,000 (or 700,000, depending on which paper you read) and that, to reduce your chances of getting cancer, you should stop eating bacon and ham.

Both these stories depend on statistics. So do many other newspaper stories. In five recent issues of the Daily Mail, I counted 19 stories that relied almost wholly on statistical data. They revealed, for example, that women treated for early signs of cervical cancer are at “double” the risk of contracting full-blown cancer 25 years later; the UK population will rise to 81 million by 2074; people who try to stop thinking about chocolate eat more of it; more than 1.2 million people have been on sickness benefits for more than five years; the sex lives of “up to” 15 million Britons are affected by stress; and eight out of 10 dog owners are “relaxed” compared with three in 10 of those who do not own pets.

I have no reason to believe these “findings” were inaccurately reported. But when I read such reports, I want to ask questions. What does “double” mean? From what to what? From one to two? What does “increased risk” mean? How big a risk? How does it compare with other risks? What does “up to” mean? In the example above, it could, taken literally, mean anything between zero and 15 million.

Journalists are not very good with figures. The great majority come from an arts or social studies background. [...] Basic statistical concepts – confidence intervals, standard deviation, probability and so on – are alien to them. Most journalism training courses do not have modules on how to handle numbers. Literacy is considered essential for reporters – or at least their subeditors – but not numeracy.

This can cause newspapers to miss good stories. Look carefully at the statistics-based stories in any newspaper and you will see that few are the result of original journalistic research. The press challenges official figures only after think-tanks, pressure groups or MPs have done the work. But perhaps more important, innumeracy leads newspapers into comical error. A couple of years ago, the Mail and the Telegraph reported that one in five men and one in eight women who reach 65 will die before 67, thereby missing out on a state pension if the retirement age were to rise to 67. The two papers were relying on a table, but had muddled the figures. The true figures were one in 29 men and one in 48 women.

That was a simple error. The reporting of medical risks raises more complex questions. It is not an error, to take a prominent example from 2005, to report that ibuprofen, a painkiller, increases the chances of a heart attack by “almost a quarter”. A paper to this effect was published in the British Medical Journal. But again, I want to ask: a quarter of what? Unless we know something about the general incidence of heart attacks, the statement is almost meaningless. In fact, the increased risk amounted to one extra heart attack among every 1,005 people taking ibuprofen.[...]

Other statistics should be treated with even greater scepticism. The Mail’s story about the benefits of owning a dog looks less convincing when you learn that the research was sponsored by a dog food brand. [...] Newspapers print this spurious nonsense more often than they should.

Much of this, you may say, is harmless fun. A story about chocolate is just a talking point. Nobody is going to change their behaviour because of it. But that is not true when newspapers report high risks in taking a painkiller. Moreover, reporting every figure as though it deserves equal credence brings statistics – an essential tool for understanding our world – into disrepute.

Statistics do not exist somewhere out in the universe as great and unquestionable truths. Somebody, somewhere, has decided that a question, worded in a particular way, should be asked, via a certain methodology and at a certain time, of a group of people defined and selected in a specific way. To understand the significance and meaning of any statistic, you need to know all of this context.

- Who commissioned and paid for the research? (They might have an interest in obtaining certain results.)
- Who did the research? (They might have an interest, or might be unqualified for the task.)
- What question was asked? (The wording or options might limit or bias answers. Alternatively, if respondents were allowed to give long, free-form answers, researchers’ interpretations might have been selective.)
- By what method was it asked? (People may respond differently to face-to-face, phone and e-mail surveys, or the method may make the sample less representative because, for example, some sections of the population don’t have phones or e-mail, or a

survey location largely populated by one type of person was chosen.)

- How was the sample population (the people who were questioned) defined and selected? (Was the selection representative of the group and large enough to have statistical significance? Bigger samples, even when not truly representative, can make more accurate predictions about the general population than smaller ones.)
- How recent or old is the research? (Maybe important context related to the question has changed since then. And over what time period was data collected?)

Bad Science

Ben Goldacre, who writes the Bad Science column in the UK *Guardian* newspaper picked up the following headline and story from another UK newspaper in his column in January 2008: “Doctors say no to abortions in their surgeries ... Family doctors are threatening a revolt against government plans to allow them to perform abortions in their surgeries ... Four out of five general practitioners (GPs) do not want to carry out terminations...” He asked questions about it similar to those above. Here’s what he discovered.

- This was an informal online poll on a doctors’ chat site, accessed by clicking the ‘abortion’ link on the site. We don’t know exactly how many doctors regularly use this particular site (or who they all are); it is one of several. We can assume that some doctors using the site self-selected in to the poll because they already had an interest in the topic – perhaps biased towards anti-abortion attitudes. And we don’t know how many doctors self-selected out by reading the survey and deciding that for whatever reason they didn’t want to fill it in.
- The question asked was “GPs should carry out abortions in their surgeries. Tick Strongly Agree; Agree; Don’t know; Disagree; Strongly Disagree.” There is no explanation of “carry out abortions in their surgeries” – in particular, no indication of whether this means with existing facilities or with extra, improved facilities, with or without extra training or staff. So the doctors responding could each have made different assumptions about the circumstances surrounding “carry out abortions” and answered accordingly.

On this basis, the figure of ‘four out of five’ GPs is meaningless, because the sample certainly doesn’t represent all GPs. And ‘threatening a revolt’ is a lie invented by the newspaper – because the poll question does not even ask GPs what they plan to do.

Basic research – skills and tools:

what are you looking at?



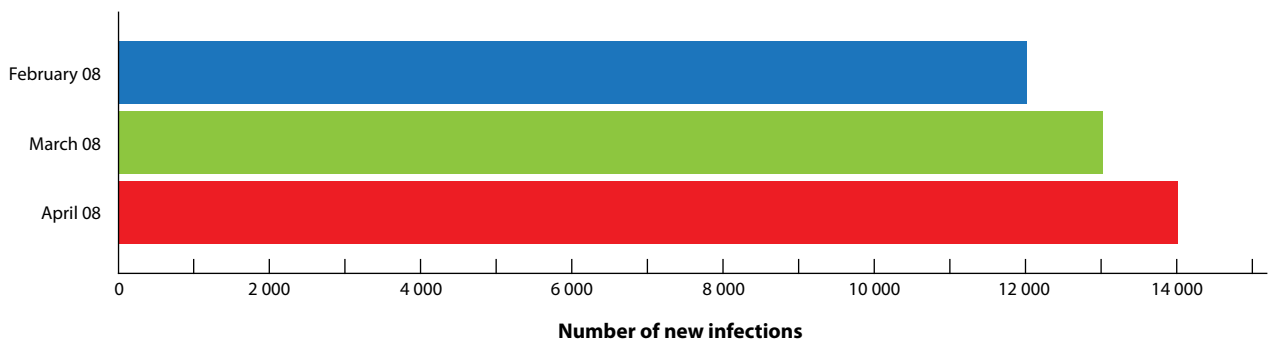
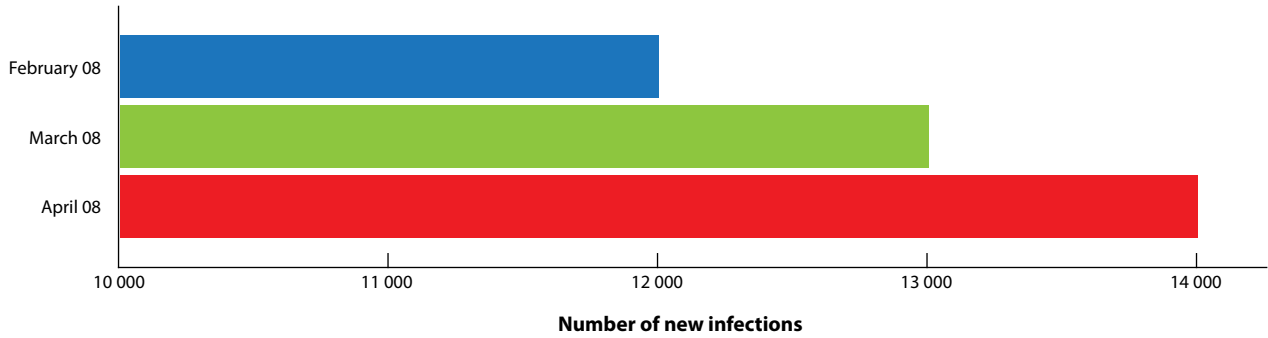
Do the results of the research indicate an absolute number, a proportion, or a rate?

Bald numbers tell you very little. You need to know the size of the whole population they are drawn from before you can say whether they are significant and how significant they are. The “four out of five doctors” figure above looks impressive. But supposing only 20 doctors filled in the survey, out of the thousands of GPs practicing in the UK. Sixteen doctors isn’t very impressive, and can’t possibly tell us anything about how thousands of doctors might think or behave.

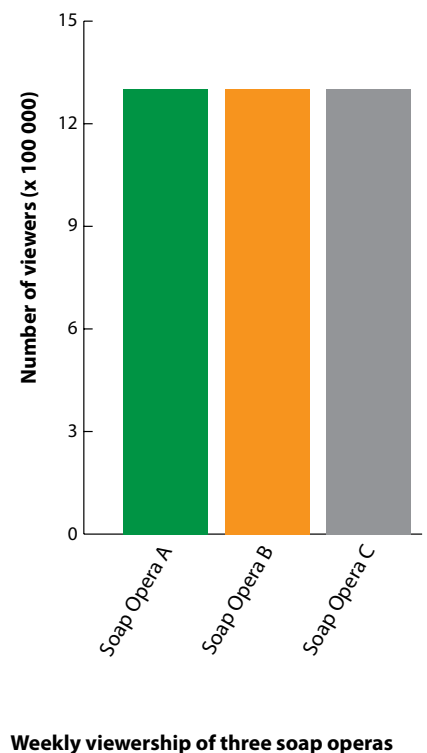
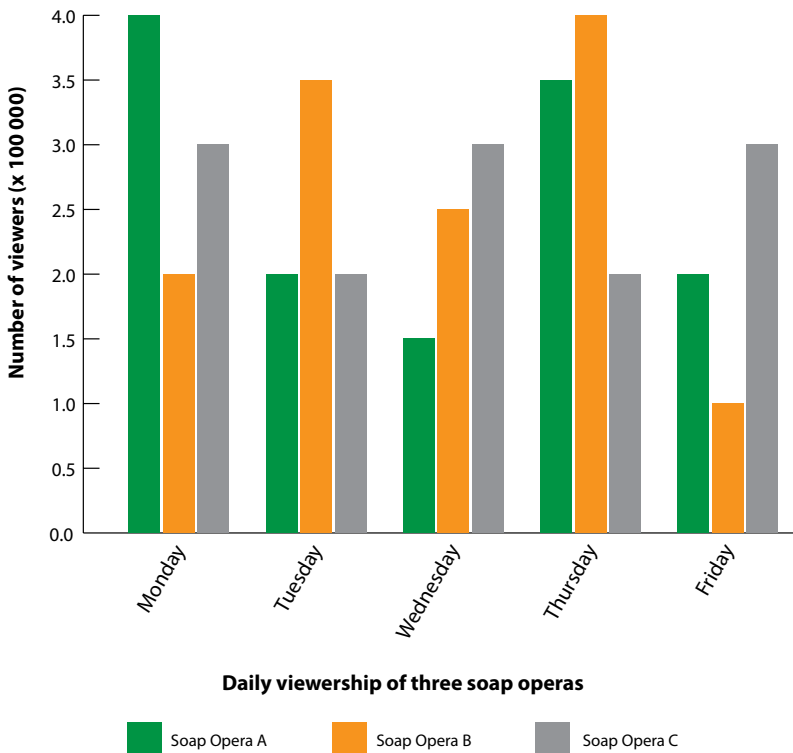
Likewise, proportions (fractions or percentages) should have us looking at the size of the whole population, and how it was defined/selected. With a rate, you need to understand the context and the meaning of the terms used, including what the direction of the change (up or down) actually means. As we’ve already explained, the ‘inflation rate’ means how fast costs are rising over time. You need to understand that, because of this definition, so long as there is a positive inflation rate, some rise in costs is happening, even if the term is attached to words like ‘falling’.

Basic research – skills and tools: **pictures and graphs**

! Always read the words! With graphs, you need to look at the scale, and the starting point. It is easy to make a tiny change look dramatic by increasing the scale and starting only with the section of the figures that show the change (see below).

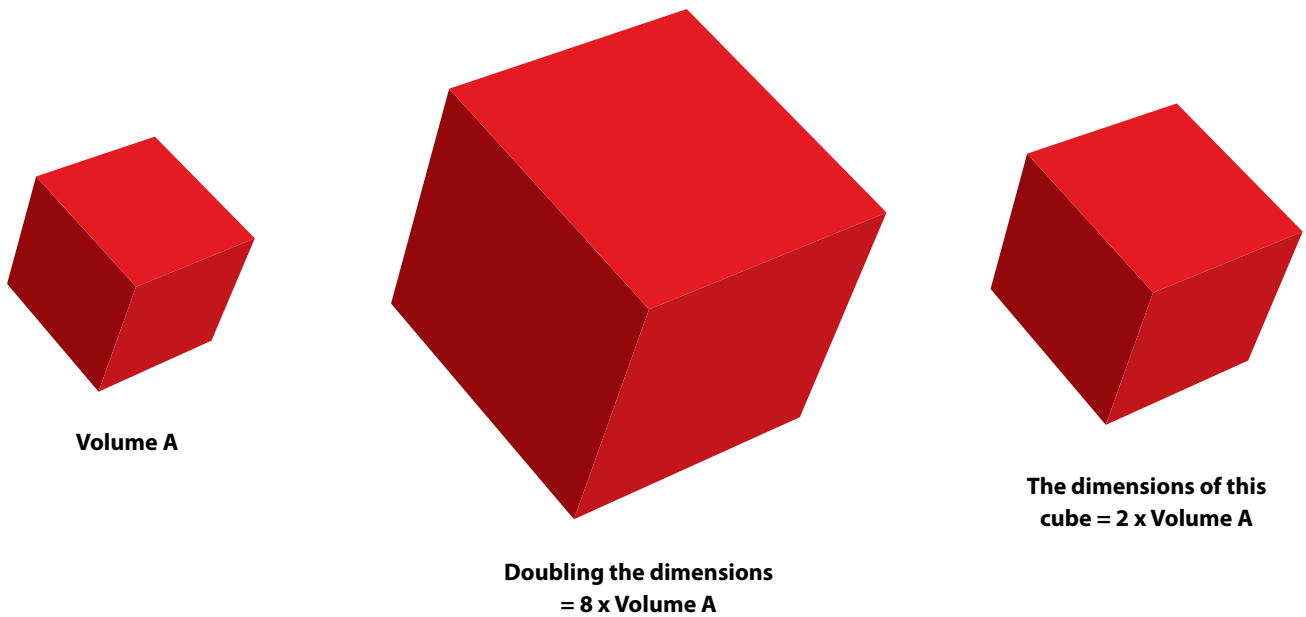


! You also get a very different picture of change from a graph showing real frequency (how often something occurred in separate time periods) and one showing cumulative frequency (adding up all the changes over time). See the charts below.





When something is represented with a symbol or in a pictogram (a chart using little drawings), it's easy to misread the real change if the artist has not respected the maths. Look at the example below:



In an attempt to indicate a doubling of the volume of something, in the central cube (above) the artist has simply increased the size of the cube two-fold. The effect of this is that the cube has been doubled in size in each dimension (length, width and height) and this means the volume is not doubled, it is **eight times bigger** ($2 \times 2 \times 2$) – which looks much more impressive. The cube on the right is in fact an accurate reflection of the doubling of the volume of the original cube.

Basic research – skills and tools: **watch the logic**

Just because two sets of figures follow the same pattern, this does not prove that they are really related to one another or that one is a cause or effect of the other. Children get bigger as they get older. Their language skills also improve as they get older, at around the same rate. But that does not mean that getting bigger improves language skills! Again, you need to read the words to find out why a relationship is being suggested. Are there precedents (valid research that has been done before on a similar or comparable area) that support the relationship being suggested?

Likewise, just because something happened after something else, this does not automatically prove that the first event caused the second. Research needs to look at the context, rule out other possible causes, and point to the precise mechanism by which the first event caused the second.

Basic research – skills and tools: **samples and what they mean**

A research sample is supposed to be representative of the bigger population the research is investigating. Supposing you were researching the attitude of parents to the latest revealing teenage fashions. First, your sample would have to include parents. But the parents of children younger than teenage may have different attitudes – and also have far more control over what their children wear. If the research included parents with children of all ages, its results might be different from a survey of the parents of teenagers.

Surveys done via telephone land-lines exclude households who do not have them. Surveys conducted in upmarket shopping malls exclude people who can't afford to shop there.

Government HIV prevalence figures in South Africa are based on the test results of mothers attending state ante-natal clinics. The statisticians then make assumptions about how these figures relate to other population groups and do projections (calculations based on mathematical assumptions) to assess the prevalence in people in categories other than expectant mothers. This gives a generally accurate broad picture, but we know from other surveys that it led in the past to an under-assessment of figures among older people, and among people from higher socio-economic groups who don't generally use state clinics. And young women who want to have babies will not, by definition, be condom users, so this group may have other characteristics that make it unique.

All of these examples illustrate that, if you are writing about research results, it's vital to find out how the original sample was defined, how results were collected, and what assumptions any projections were based on.

Figures are not neutral

Reuters recently published a story by journalist Ruth Gidley on how humanitarian statistics are collected. These extracts show just how human – and controversial – the business of collecting statistics is. Find the full story on the Reuters Website.

LONDON – To count the dead they ride motorbikes, charter planes and wade through snake-infested rivers. The precious statistic can help aid agencies convince a weary world there is a crisis in the jungles of Africa or forgotten corners of Iraq – where death comes from hunger and disease related to war as well as from war itself ... But pinning down precise, and credible, numbers is difficult and often intensely political. To prove its point, the IRC and other agencies have to go to where the dead are [...]

Their Congo surveys haven't been particularly controversial, but researchers trying to pin down a death toll for Iraq know their findings will come under intense scrutiny.

"People who oppose the war usually cite the highest death toll they can find, and people in favour of it tend to cite the lowest," said John Sloboda, who co-founded Iraq Body Count to collate reports from the media and mortuaries. The British-based Web site – set up in 2003 before the start of the U.S.-led invasion – seeks to commemorate those who have died with entries that give the person's name whenever possible. "It's just an act of humanity to record and memorialise the dead," Sloboda said. "We've done it for soldiers for centuries." [...]

The Iraqi Health Ministry and the World Health Organisation said this year they calculated 151 000 violent deaths between 2003 and 2006. A team from Johns Hopkins University in the United States estimated as many as 654 000 deaths beyond the norm in Iraq during the same period. A much simpler survey by British pollsters Opinion Research Business (ORB) put Iraqis killed by the war at approximately 1.03 million, based on a finding that 18 percent of the 2 414 adults they interviewed cited at least one death in the household due to the war. ORB calculated its number using Iraq's last census in 1997, which documented 4.05 million households.

The International Red Cross reports highlight the fact hunger and disease caused by a conflict often kill more people than the violence itself. In Congo, for instance, intentional deaths count for less than 10 percent of the total. Most people die when they are cut off from food supplies or medical facilities, or because the health system has been nearly eliminated by the conflict. The IRC findings from Congo appear to mirror experiences in Angola, Liberia, Sierra Leone; all still struggling to recover from brutal conflicts.

(For more information on humanitarian issues visit www.alertnet.org)

Basic research – skills and tools:

look for definitions and numbers

Are loose general terms such as 'positive developments' used in reporting the research? Are they explained or defined? What does the researcher mean by terms like 'possibility', 'probability', 'likely' or 'unlikely'? There are mathematical ways of tying down probability, and even providing precise odds for how likely some event or causation is. But without these, you have to look for a relationship between the analysis and the figures given, and check whether the researcher has been consistent in how they use terms.

The key question to ask about statistical and numerical research is always: could the same figures be interpreted to mean something else, or processed differently (i.e. by changing the categories or definitions) to come up with different results? If you don't have the know-how to determine this, you need a tame numbers expert in your contacts book who can help you.

Basic research – skills and tools:

remember that data alone can't prove anything

Figures are very rarely 'proof' of anything. As we have seen, the assumptions on which research is based and interpreted can create very different 'spin' on the bald numbers.

- Data detailing only which NGOs in your country received donor aid do not tell us anything about how well or badly that aid was used to achieve its goals. You could interpret it to support a call for aid to be increased or to be cut, depending on which examples in the list you selected.
- Data proving that a food additive does not cause cancer do not remove the possibility that it may cause other illnesses. Nor do the data provide evidence that the additive does good, or is necessary in food production.

Figures, or course, are only one sort of data. They are called ‘hard’ data, and this implies that they are more solid and reliable. They often are. But ‘soft’ data – research into attitudes and other social aspects, may provide more useful information on some topics. And while expert knowledge is vital, local knowledge (from, for example, old people in your community who have watched developments over years) can also provide valid insights. When the media consistently privilege hard over soft and expert over local, they weaken their own role in democracy. As a journalist, you should be working with all types of data from all sources, using skilful contextualisation to help your readers weigh up the merits of each.

Data in areas of uncertainty – perhaps because the field of study is new, and controversies have not yet been resolved – are the hardest for journalists to deal with. Interested parties can exploit journalists’ ignorance in such areas by making the topic sound even more uncertain. This was the strategy used by the tobacco industry worldwide to exploit the media’s lack of specialist knowledge. Before the link between cigarette smoking and lung cancer had been definitively proved, they paid researchers to generate alternative research, making it look as though the argument about the role of cigarettes in cancer was evenly balanced. The energy industry has done the same thing very successfully in the debate about global warming. The weight of peer-reviewed, credible scientific research has for a long time been that human activity has changed the climate and the environment in damaging ways. But any journalist simply Googling the topic will find an equal number of articles arguing the opposite, many of them generated by impressive-sounding ‘research institutes’ financed by the energy industry.

Finally, research is often used by those in power to limit policy options and debates. Suppose the government finances a research project to discover “Should we pay people individual compensation for a new dam which will flood their farmlands, or would it be better to pay for a relocation scheme that resettles everybody?” It doesn’t matter how ‘hard’ the data are in favour of one option, key questions about other options – including “should we build the dam at all?” – have not been asked. So don’t feel you must always limit your story to the terms of other people’s research questions.

Remember the following

- Many specialist areas are still being explored. You need an expert (preferably more than one, and from different sides of the debate) to help you read through the spin of what sounds like a definitive answer, even if it has hard figures attached.
- How people collect and interpret data may be influenced by context, including social, political and cultural factors, the history of debates in the area, and who is funding the research.

Exercise #6

Spotting the story numbers

In a press release, your government claims that measures to restrict cheap Chinese clothing imports have been successful and have boosted growth in the local and neighbouring African clothing industries. It cites the following figures – but what do they REALLY tell you? (And what do they conceal?)

CLOTHING IMPORTS TO OUR COUNTRY (\$m value)

Country of origin	2006 (before import legislation)	2007 (after import legislation)
China	980	700
Rest of Africa	80	220
Other (main countries: India, Bangladesh, Myanmar, Vietnam)	20	60

- Did you notice the quantity of imports hasn’t changed – which suggests your own country’s clothing industry is still facing tough competition from imports.
- Imports from the rest of Africa have nearly trebled. But are these from African-owned factories, or from overseas-owned factories based in Africa?
- Imports from other non-African countries have trebled. Two of those countries (Vietnam and Myanmar) are in the Far East, and there is significant Chinese ownership of factories in Vietnam. In addition, Myanmar is a country currently ruled by a repressive military dictatorship, and many campaigners have suggested that buying their exports is actually providing support for the regime.

So where, really, is the story? In the official figures, or somewhere else?

Case studies

by Dr Joseph Hanlon

This case study illustrates a range of important points about research technique and the use of figures. Hanlon's jumping-off point was a glowing World Bank press release – which did not actually contain any useful figures. Hanlon had to obtain the precise figures and do the calculations in order to prove that the press release was misleading. The investigation was published as a series of press releases and web articles 23 April, 5 May, 6 May and 4 June 1998. The main article was published in *Metical* on 22 April 1998, but there were various follow up articles and IMF replies. It had a massive impact. Both on Mozambique's debt burden, and on the policy of international financial institutions.

The International Monetary Fund (IMF) and the World Bank have immense power over poor countries and they seem unchallengeable because they claim to have the best and most highly paid economists in the world. How can an individual investigative journalist challenge that power? But the media has a place keeping even the international financial institutions in check.

A story that I wrote showed the World Bank to be dishonest, and helped to force it cancel some of Mozambique's debts. That research showed three key points:

- Absolute power breeds arrogance and sloppiness in all institutions, including the IMF and World Bank, and leads to gross errors which can be discovered by investigative journalists.
- Spreadsheets, publicly available data, and patience can be more important tools for the investigative journalist than secret sources.
- Pressure groups can be an important route for investigative journalism – their special interests sometimes mean there is more time and space to do the necessary research.

I am based partly in Maputo, Mozambique and partly in London, England, largely as a free-lance journalist. But I have also used my investigative journalism skills in other ways, directing the Commonwealth Independent Expert Study on Sanctions Against South Africa in the late 1980s, as a policy officer for the Jubilee 2000 campaign to cancel poor country debt in the late 1990s, and more recently as a Senior Lecturer in development studies at the Open University in the UK. But in parallel to my work in these posts, I have always continued to write books and articles, and a decade ago I was writing regularly for the Maputo faxed daily *Metical*.

A decade ago the issue of unpayable poor country debt was just coming to prominence. The Jubilee 2000 campaign was promising a massive demonstration at the meeting of the leaders of the G8 – the group of the eight richest countries – in Birmingham, England, 16 May 1998. (50,000 people formed a human chain around the conference venue to symbolise the chains of debt.) The World Bank and IMF were anxious to show they were doing something, and on 7 April 1998 announced the Heavily Indebted Poor Countries (HIPC) Initiative. Mozambique was one of the poorest countries in the world and was spending more on debt service payments than on health and education, so it was a priority country. In a press statement on 7 April 1998 the World Bank said that the HIPC debt "relief" would "free budgetary resources and allow Mozambique to broaden the scope of its development effort." In other words, money that had been used to repay debt could instead be used for health and education. The only problem was that the statement was false.

I began my investigation because the World Bank press release did not "feel" right. At that time, the whole HIPC process was surrounded by secrecy and confusion, and the World Bank press release (<http://go.worldbank.org/6QYFWXCP60>) gave no useful figures. But through another campaign group I was able to obtain a copy of the 31 March 1998 HIPC Final Document which had been approved by both IMF and World Bank boards. (Such documents are now public, making what I did then easier now, but no less essential.) The first shock was that the Bank and Fund staff had also not told their own directors how much Mozambique's payments were reduced, and had not even provided sufficient information for the directors to calculate the figure. All the directors received was a bland statement similar to the press release; they approved the package because staff told them it was good, but they had no way to make an informed judgement.

I had recently written a book on Mozambique development, called *Peace Without Profit: How the IMF Blocks Rebuilding in Mozambique* (1996, Oxford: James Currey) and I had obtained quite a range of IMF and World Bank documents from sources in Mozambique, both in government and in the donor community. Putting data from the HIPC Final Document and from one of the other IMF documents into a spreadsheet, I was able to actually calculate how much Mozambique had been paying and how much it was going to pay after HIPC debt "relief". And the answer was \$107 million dollars a year before HIPC, and slightly more after HIPC debt "relief".

In other words, World Bank and IMF staff had not only been dishonest in a public statement, they had misled their own boards. I wrote an article for *Metical* published 22 April 1998 and issued a press release for Jubilee 2000 on 23 April 1998. At first, the IMF disputed my calculation. Then, just days before the Birmingham G8 meeting and demonstration, on 12 and 14 May, the IMF took the unprecedented step of releasing two different sets of figures on Mozambique's debt service payments. The two sets of figures disagreed – but both sets showed that my figures and press release had been right. Under HIPC, the Bank and Fund were only cancelling debt which Mozambique was not repaying anyway, and the Bank press release had been wrong and misleading.

What was the outcome?

The G8 in Birmingham accepted that HIPC was not sufficient, and called for further debt cancellation. A year later, on 2 April 1999, Bank and Fund staff admitted to their boards that I had been right: debt service payments after HIPC debt relief “are not dramatically different” than before because HIPC only cancelled debt that was not being repaid and never would be repaid. Then, on 30 June, the Bank and Fund recalculated and made a real cut in Mozambique’s debt service payments – to \$70 million a year – and later made another cut to \$50 million a year. This did mean that \$1 million a week was made available for additional health and education spending in Mozambique.

I have done many investigative stories, including work on the murder of a friend, *Metical* editor Carlos Cardoso. But I have chosen this story, even though it is older, because I think it had more impact than any other article I have ever written – it proved the World Bank and IMF were lying about debt relief. Of course, one story never changes the world, but this one helped to change the mood of the G8 leadership, helped to push for a change in Bank and Fund policy, and helped to save Mozambique \$1 million per week – not bad for a single investigation.

I have also chosen this story because it illustrates a series of key points about investigative journalism:

- The importance of specialisation. I had worked on books about Mozambique and had substantial economic documentation, including IMF reports. I then began to specialise in debt, giving me other contacts and access to other reports. My investigations usually start from that intangible sense of “feel” – something in the back of my head told me the World Bank was wrong, and that came purely from experience of having worked on Mozambique and debt. Specialisation also gave me access to documents.
- The middle ground between public and secret is often most useful for journalists. IMF and World Bank reports, for example, are often supposedly “confidential” but they are distributed to so many people that it is easy to get copies. Many more things are now on the web.
- Investigation can often just be a lot of reading, a lot of patience, and some experimentation with a spreadsheet. This investigation had no secret sources. Rather it involved pulling together material from a range of different documents and using it in a way that World Bank and IMF staff had not expected – part of the arrogance of power is the assumption that if information is not handed out on a plate, journalists will not bother to look elsewhere.
- The importance of accuracy and trustworthiness. My press release for Jubilee 2000 was taken seriously by economic journalists and by the IMF and British Treasury because we had established a track record of always being able to prove the statements we made. One error or exaggeration can destroy the reputation and credibility of an investigative journalist. But the other side is that a journalist with a good record is trusted by both sources and readers.
- Investigative journalism often takes place outside the mainstream media. I used these techniques for Jubilee 2000 press releases and for research leading to academic articles published in prestigious journals. That also means that reporters should never ignore pressure groups and academics, because they have the time and motivation to do serious investigation. Of course, they often have vested interests – as I did when I worked for Jubilee 2000 – but if the results can be backed up, they can be explosive.

Investigative journalism is hard work. But it can bring about real changes.

Key points from this chapter

- ✓ Use data management tools – either via computer software or a good secure filing system – to build up your paper trail and keep the detail of your investigation under control.
 - ✓ Develop your CAR, profiling, paper trailing and data mining skills.
 - ✓ If your country has Access to Information laws – use them.
 - ✓ If your country does not yet have such laws, join the campaigns to secure them.
 - ✓ Remember that by working with and through organisations in countries that do have such laws, you may be able to create a ‘back door’ to access information.
 - ✓ Look for ways you can use numbers and statistics to strengthen even social and ideas-based stories.
 - ✓ Look into numerical data to spot story ideas and angles.
 - ✓ Master basic number skills.
 - ✓ Always interrogate numbers and statistics to discover their source and how they were compiled.
-

- ✓ Always re-check figures – your own, and those supplied by others – to ensure that they have been calculated correctly.
- ✓ Remember statistics are compiled by human beings; they are not unquestionable and very rarely provide ‘proof’ on their own.
- ✓ Ask questions about numbers by means of short, closed, staged questions to ensure you get precise answers.
- ✓ Make numbers accessible for readers by rounding them or explaining them, but do not distort them when you put them into words.

Glossary

- **Access to information/Freedom of information** – (The need for) legislation that allows the public access to government and private sector records that are in the public interest
- **Computer Assisted Reporting** – using the internet to find and analyse information
- **Chronology building** – placing events found from records on a timeline in order to find what happened (and who was where) at any given time of the (series of) events you are researching
- **Database and database management** – a set of statistics, recorded or found facts that are kept by an institution that manages it to make findings and keep track of developments. A journalist can establish and manage his/her own database
- **Data-mining** – searching databases methodically in order to make findings
- **Paper trail** – the trail that appears when one record leads to another
- **Parallel backgrounding** – using records from different entities and comparing them to make findings about one particular event or individual
- **SAHA** – the South African History Archives, who use the South African Access to Information legislation in order to access SA government and private sector records

Further reading

- Read Stiglitz and Bilmes’ investigation of the costs of the Iraq war in their book *The Three Trillion Dollar War* (Norton 2008) or read a short summary of their work at http://www.timesonline.co.uk/tol/comment/columnists/guest_contributors/article3419840.ece, Feb 2008.
- For more information on database management and different tools with which to map out statistics and databases, check www.ire.org and www.nicar.org. These sites are also useful for tutorials and free tips on various aspects of research techniques
- The UK CIJ has a good links page for search engines and useful websites and databases: <http://www.tcij.org/links>
- The FAIR website www.fairreporters.org provides (under ‘resource centre’, ‘library’ and ‘links’) a list of websites that assist with on line research (like encyclopedia’s, translation services and a search engine for old web pages)
- For the current state of media legislation in African countries: KAS has published, in PDF, on its website, outlines of media law in Mozambique, the DRC, South Africa, Malawi, Botswana and Namibia.
- Find a list of organisations involved in the access to information struggle, as well as a list of specialised websites, at the end of Chapter 5.
- The best book on understanding statistics, although now more than half a century old, remains Darryl Huff’s *How To Lie With Statistics* (1954).
- Good on line profiling tools are the SA company registry (accessible at www.sacompany.co.za, at an affordable subscription rate, or through the FAIR Helpdesk; www.deedsearch.co.za; and Google Earth at <http://earth.google.com>. The FAIR Helpdesk also offers assistance with searches of company registries internationally.
- An excellent book ‘Numbers in the Newsroom’, by Sarah Cohen, can be ordered from IRE (www.ire.org).