

Reproducibilidad en Investigación

Sesiones de estadística DAP-Cat

Miguel A. Martinez-Beneito
miguel.a.martinez@uv.es

Reproducibilidad: el concepto

- ▶ **Reproducibilidad:** Cualidad de un experimento de poder ser **repetido** en cualquier lugar y por cualquier persona.

WIKIPEDIA La enciclopedia libre

Página Principal Portada de la comunidad Actualidad Cambios recientes Páginas nuevas Página aleatoria Ayuda Documentación Notificar un error Imprimir/exportar Crear un libro Descargar como PDF Versión para imprimir

En otros proyectos Wikimediacommons Wikisource Herramientas Lo que enlaza aquí Cambios en enlazados Subir archivo Páginas especiales Enlace permanente Información de la página Elemento de Wikidata Citar esta página En otros idiomas Afrikans

Método científico

Este artículo trata de los diversos métodos científicos utilizados históricamente. Para el uso de la palabra en entornos técnicos actuales ver [investigación científica](#).

El **método científico** (del griego: μετά, metá 'hacia, a lo largo οδός, hodos' 'camino'; y del latín: *scientia* 'conocimiento'; 'camino hacia el conocimiento') es un método de investigación usado principalmente en la producción de conocimiento en las ciencias. Para ser llamado científico, un método de investigación debe basarse en el empírico y en la medición, sujeto a los principios específicos de las pruebas de razonamiento.¹ Según el *Oxford English Dictionary*, el método científico es: «un método o procedimiento que ha caracterizado a la ciencia natural desde el siglo XVII, que consiste en la observación sistemática, medición, experimentación, la formulación, análisis y modificación de las hipótesis».²

El método científico está sustentado por dos pilares fundamentales. El primero de ellos es la **reproducibilidad**, es decir, la capacidad de repetir un determinado experimento, en cualquier lugar y por cualquier persona. Este pilar se basa, esencialmente, en la comunicación y publicidad de los resultados obtenidos (por ej. en forma de [artículo científico](#)). El segundo pilar es la **refutabilidad**, es decir, que toda proposición científica tiene que ser susceptible de ser falsada o refutada ([falsacionismo](#)). Esto implica que se podrían diseñar



René Descartes, filósofo, matemático y físico francés, considerado el padre de la filosofía moderna, así como uno de los nombres más destacados de la revolución científica. [62]

- ▶ Reproducibilidad = **transparencia**.
- ▶ La reproducibilidad da **carácter empírico** al método científico y permite **constatar** y validar pruebas **experimentales**.

- ▶ Sin reproducibilidad, la ciencia se reduce a **confianza mutua** entre experimentadores. Por tanto la reproducibilidad distingue la **ciencia de la fe**.

'A Newton o a Galileo no les bastaba con hacer descubrimientos capaces de cambiar la historia. **Debían además repetir sus experimentos delante de todos sus colegas**, y estos, a su vez, los repetían por su cuenta antes de quedar completamente convencidos. Este principio de reproducibilidad ha sido fundamental para el avance de la ciencia desde entonces. En la actualidad, **esa garantía esencial se está perdiendo**, y pone en entredicho la validez de muchos estudios en casi todas las disciplinas.

Al final, los científicos se creen lo que ven, pero no hay forma de comprobar que es cierto, y además no podemos usar esos datos posteriormente porque se han esfumado. **Esta falta de transparencia es uno de los mayores retos que afronta la ciencia.'**

El País, 10 enero 2017

Estadística forense y reproducibilidad: como muestra un botón.

The Annals of Applied Statistics
2009, Vol. 3, No. 4, 1309–1334
DOI: 10.1214/09-AOAS291
© Institute of Mathematical Statistics, 2009

DERIVING CHEMOSENSITIVITY FROM CELL LINES: FORENSIC BIOINFORMATICS AND REPRODUCIBLE RESEARCH IN HIGH-THROUGHPUT BIOLOGY

BY KEITH A. BAGGERLY¹ AND KEVIN R. COOMBES²

[Annals of Applied Statistics \(2009\)](#)

- ▶ **Enviado** a evaluar/publicar el 1 sep. 2009. **Publicado** online el 14 sep. 2009.
- ▶ Discute **reproducibilidad** como problemática en general y lo **ilustra** con lo encontrado por los autores al intentar reproducir **varios estudios**.
- ▶ Nos centraremos en el **primero** y **tercero** de sus 4 **case studies**.

Antecedentes: Potti et al. (2006)

Genomic signatures to guide the use of chemotherapeutics

Anil Potti^{1,2}, Holly K Dressman^{1,3}, Andrea Bild^{1,3}, Richard F Riedel^{1,2}, Gina Chan⁴, Robyn Sayer⁴,
Janiel Cragun⁴, Hope Cottrill⁴, Michael J Kelley², Rebecca Petersen⁵, David Harpole⁵, Jeffrey Marks⁵,
Andrew Berchuck^{1,6}, Geoffrey S Ginsburg^{1,2}, Phillip Febbo¹⁻³, Johnathan Lancaster⁴ &
Joseph R Nevins¹⁻³

Nature Medicine, 2006

- ▶ Una de “the **top 6** genetics histories of 2006” ([Discover, 2007](#))
- ▶ Gran número de **citas** en muy poco tiempo.
- ▶ Investigadores de **MD Anderson interesados** en desarrollar estudios similares.

El interés de Potti et al. (2006)

- ▶ **No todos** los pacientes con tumores en una misma localización son **sensibles** a un mismo tratamiento.
- ▶ **Determinación** previa de **sensibilidad** a tratamientos permitiría **tratamiento personalizado** para cada individuo, y éste, a su vez:
 - ▶ Evitar **efectos adversos** sobre un gran número de pacientes.
 - ▶ Disminución de **costes** sanitarios.
- ▶ **Objetivo:** Determinación **sensibilidad** a tratamientos basada en el **perfil genético** de cada tumor.
- ▶ **Novedad:** No determinan **un único gen** sino la combinación (mediante machine learning) de un **conjunto** de genes para predecir sensibilidad/resistencia a tratamientos.

Case study I: Adriamicina

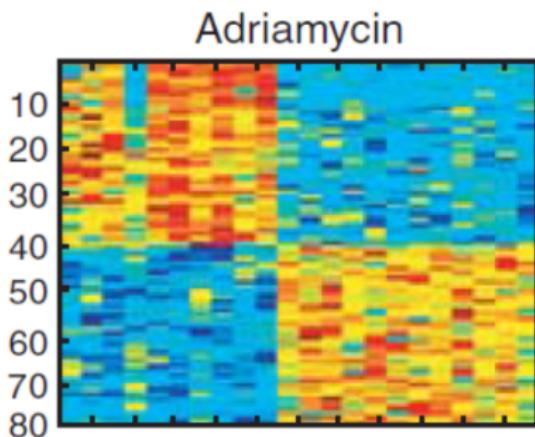
- ▶ Distintas muestras de **células cancerosas** se someten a tratamiento de **quimioterapia** (Adriamicina).

Diseño:

- ▶ **144** muestras: 22 training, 122 testing.
- ▶ Datos provienen de **estudios previos** (afortunadamente públicos):
 - ▶ Variable respuesta: **sensibilidad/resistencia** a Adriamicina.
 - ▶ Variables explicativas: Microarray con **8959** valores del **perfil genético** de cada muestra.
- ▶ **Objetivo** de Baggerly: **Validación de capacidad predictiva** descrita por Potti para grupo “testing”.

Case study I: Resistant vs sensitive

- ▶ Grupo training: “[10] Resistant lines on the left and [12] sensitive on the right” (Potti et al., 2006)

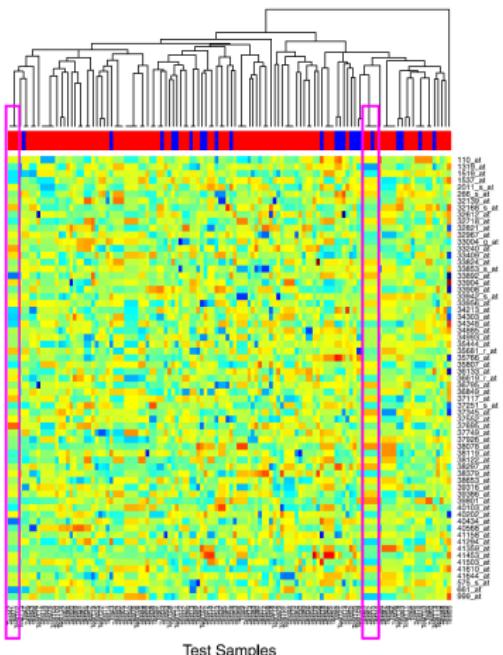


- ▶ Baggerly replican el análisis de Potti.
- ▶ Encuentran en el grupo training (datos originales): **12 tumores resistentes, 10 tumores sensibles.**
- ▶ !!Los resultados reales van en **sentido opuesto** a lo publicado!!

- ▶ Aun así continuan con la **validación** de los resultados de Potti publicados para el **grupo testing**.

Case study I: Validación testing data

- ▶ Validación de los resultados para el grupo testing por Baggerly.



- ▶ Los **80 genes** en este gráfico corresponden a aquellos determinados por Potti en su análisis para el grupo training.
- ▶ Dicho análisis muestra datos **sospechosamente parecidos**
- ▶ !!Hay muestras repetidas!! sólo **84** de las **122** muestras son **distintas**, hay datos repetidos hasta 4 veces.
- ▶ Hay muestras repetidas que son a su vez **sensibles y resistentes**.

- ▶ Pero, independientemente de los repetidos (y que los resultados puedan ir en sentido contrario) ¿**Por qué los genes** determinados por Potti **no discriminan** los tumores del grupo testing?
- ▶ Baggerly reproduce el análisis de Potti y encuentra la siguiente **lista de genes** como potenciales **discriminantes** de la sensibilidad a Adriamicina:

| Theirs | Ours |
|-------------------|--------------|
| ... | |
| [3,] "1881_at" | "1882_g_at" |
| [4,] "31321_at" | "31322_at" |
| [5,] "31725_s_at" | "31726_at" |
| [6,] "32307_r_at" | "32308_r_at" |
| ... | |

- ▶ Las **etiquetas** de los genes están **movidas** una posición.
- ▶ En resumen, el **trabajo de Potti** no es más que el producto de una **concatenación de errores** y sus resultados **carecen de** ningún **valor**.

Case study III: Otra vez Potti . . .

- ▶ Potti et al. (2006) predicen, además, el éxito de **combinaciones de tratamientos** en función de las **probabilidades individuales** de éxito de cada tratamiento (ejemplo de Adriamicina).

Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial

Hervé Bonnefoi, Anil Potti, Mauro Delorenzi, Louis Mauriac, Mario Campone, Michèle Tubiana-Hulin, Thierry Petit, Philippe Rouanet, Jacek Jassem, Emmanuel Blot, Véronique Becette, Pierre Farmer, Sylvie André, Chaitanya R Acharya, Sayan Mukherjee, David Cameron, Jonas Bergh, Joseph R Nevins, Richard D Iggo

[Lancet Oncology \(2007\)](#)

- ▶ Bonnefoi et al. **validan** la propuesta de Potti et al. (2006)
- ▶ Baggerly se plantea la **validación** de los resultados de **Bonnefoi**.

Case study III: leyes probabilidad

Potti et al. (2006)

Methods. In instances where a combined probability of sensitivity to a combination chemotherapeutic regimen was required based on the individual drug sensitivity patterns, we used the probabilities of response to individual drugs and used the theorem for combined probabilities as described by William Feller to deduce a probability of response to a combination of the drugs being studied. The result was then mean-centered to give a probability between 0 and 1. Hierarchical clustering of tumor predictions was performed

- ▶ Feller, W. (1968). An introduction to probability theory and its applications: volume I. New York: John Wiley & Sons. Página 22

Theorem. For any two events A_1 and A_2 the probability that either A_1 or A_2 or both occur is given by

$$(7.4) \quad \mathbf{P}\{A_1 \cup A_2\} = \mathbf{P}\{A_1\} + \mathbf{P}\{A_2\} - \mathbf{P}\{A_1 A_2\}.$$

- ▶ O, de forma más general:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- ▶ Lo que **uno se encuentra** en Potti et al. (2006) recurriendo a “estadística forense” . . . :

$$P(T \cup F \cup A \cup C) = P(T) + P(F) + P(A) + P(C) - P(T)P(F)P(A)P(C)$$

donde T=taxol, F=Fluororacil, A=Adriamicina, C=Ciclofosfamida.

- ▶ y !!**ajustan probabilidad** más alta a 1, la más baja a 0!! e interpolan el resto.
- ▶ Pero en Boneffoi et al. tampoco les falta imaginación.
“Reglas” usadas para otras combinaciones de tratamientos:

$$P(T \cup E \cup T) = P(T \cup E) = \max(P(E), P(T))$$

donde $T \cup E \cup T$ denota: Taxotere+ Epirubicina+ Taxotere (2^a sesión), o:

$$P(F \cup E \cup C) = \frac{5}{8}(P(F) + P(E) + P(C)) - \frac{1}{4}$$

Reproducibilidad y protección de datos

UKCTOCS (UK Collaborative Trial of Ovarian Cancer Screening): **ensayo clínico** aleatorizado con 200,000 mujeres que valora la efectividad de un programa de screening de **cáncer de ovario**.

Primeros resultados del programa publicados a finales de 2015, Lancet (2016). Seguimiento de 14 años.

Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial



Ian J Jacobs*, Usha Menon*, Andy Ryan, Aleksandra Gentry-Maharaj, Matthew Burnell, Jatinderpal K Kalsi, Nazar N Amso, Sophia Apostolidou, Elizabeth Benjamin, Derek Cruickshank, Danielle N Crump, Susan K Davies, Anne Dawney, Stephen Dobbs, Gwendolen Fletcher, Jeremy Ford, Keith Godfrey, Richard Gunu, Mariam Habib, Rachel Hallett, Jonathan Herod, Howard Jenkins, Chloe Karpinskyj, Simon Leeson, Sara J Lewis, William R Liston, Alberto Lopes, Tim Mould, John Murdoch, David Oram, Dustin J Rabideau, Karina Reynolds, Ian Scott, Mourad W Self, Aarti Sharma, Naveena Singh, Julie Taylor, Fiona Warburton, Martin Widschwendter, Karin Williamson, Robert Woolas, Lesley Fallowfield, Alistair J McGuire, Stuart Campbell, Mahesh Parmar†, Steven J Skates†

Summary

Background Ovarian cancer has a poor prognosis, with just 40% of patients surviving 5 years. We designed this trial to establish the effect of early detection by screening on ovarian cancer mortality.

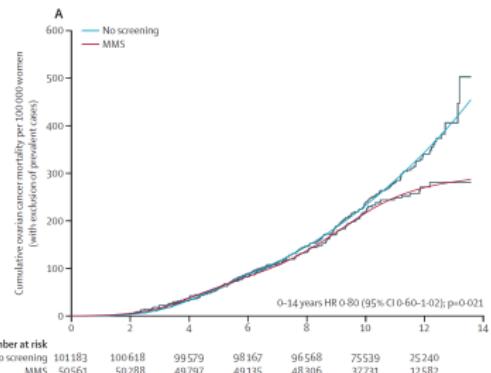
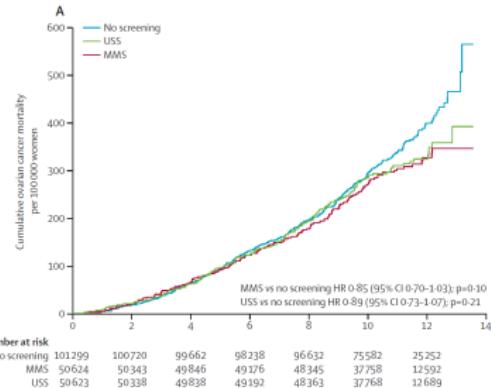
Lancet 2016; 387: 945-56

Published Online
December 17, 2015

¿Incrementa la supervivencia el programa de screening?

Resultados UKCTOCS:

Resultados mixtos. El análisis original (preregistrado) no da resultados significativos, pero sí un **reanálisis** (post-hoc) que se argumenta más apropiado.



Log-Rank test P=0.10.

Post hoc weighted Log-Rank test P=0.02.

| | Number of women (n) | Deaths (n) | Mortality reduction 0-14 years (%) | p value | Mortality reduction 0-7 years (%) | Mortality reduction 7-14 years (%) |
|--|---------------------------|---------------|--|---------|---|--|
| Ovarian cancer (primary analysis) | | | | | | |
| Cox model | | | | | | |
| MMS | 50624 | 148 | 15% (-3 to 30) | 0.10 | .. | .. |
| USS | 50623 | 154 | 11% (-7 to 27) | 0.21 | .. | .. |
| No screening | 101299 | 347 | .. | .. | .. | .. |
| Royston-Parmar model | | | | | | |
| MMS | 50624 | 148 | 16% (-1 to 33) | 0.11 | 8% (-20 to 31) | 23% (1 to 46) |
| USS | 50623 | 154 | 12% (-6 to 29) | 0.18 | 2% (-27 to 26) | 21% (-2 to 42) |
| No screening | 101299 | 347 | .. | .. | .. | .. |
| Royston-Parmar model (excluding prevalent cases) | | | | | | |
| MMS | 50561 | 120 | 20% (-2 to 40) | 0.021 | 8% (-27 to 43) | 28% (-3 to 49) |
| No screening | 101183 | 281 | .. | .. | .. | .. |
| Weighted log-rank (post-hoc) | | | | | | |
| MMS | 50624 | 148 | 22% (3 to 38)* | 0.023 | .. | .. |
| USS | 50623 | 154 | 20% (0 to 35)* | 0.049 | .. | .. |
| No screening | 101299 | 347 | .. | .. | .. | .. |

Resultados significativos ($p<0.05$) !!pero **IC contiene al 0!!**

Baggerly decide **pedir datos** (Periodo de seguimiento, Tratamiento, Outcome (viva/muerta)) para contrastar resultados.

UKCTOCS **se niega a compartir** datos “para preservar confidencialidad”.

Análisis paralelo

Baggerly decide hacer un **análisis paralelo**.

Guyot et al. BMC Medical Research Methodology 2012, 12:9
<http://www.biomedcentral.com/1471-2288/12/9>



TECHNICAL ADVANCE

Open Access

Enhanced secondary analysis of survival data:
reconstructing the data from published
Kaplan-Meier survival curves

Patricia Guyot^{1,2}, AE Ades¹, Mario JNM Ouwens³ and Nicky J Welton¹

Liu et al. Systematic Reviews 2015, 3:151
<http://www.systematicreviewsjournal.com/content/3/1/151>



Open Access

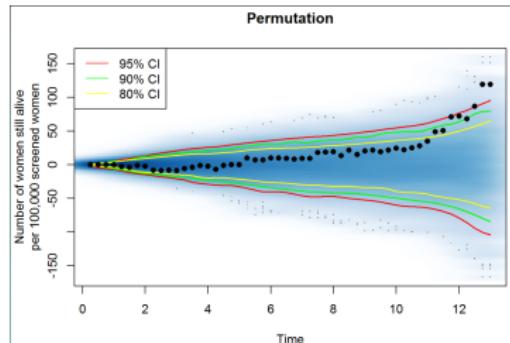
METHODOLOGY

Recovering the raw data behind a non-parametric
survival curve

Zhihui Liu^{1,2}, Benjamin Rich³ and James A Hanley^{4*}

Curvas **Kaplan-Meier** en formato pdf (vectorial) contienen la
información requerida.

Se extrae información de los
pdf de los artículos y,
finalmente, se puede hacer un
análisis paralelo.



Algunas reflexiones sobre reproducibilidad

► El caso de Potti es un **caso extremo** de malas prácticas:

- 10 artículos **retractados** (6 más parcialmente).
- **Expulsado** de su Universidad.
- **Denunciado** por pacientes de ensayos clínicos.
- Página en **Wikipedia** contando sus andanzas.
- Espacio monográfico en **60 minutos**.

Anil Potti's Wikipedia page details his academic background and research interests. His biography notes he graduated from Christian Medical College, Vellore, India in 1985 and received an MD from the University of North Dakota School of Medicine in 1989. He worked at Duke University from 2000 to 2010, during which time he was involved in a cancer research project. The page also lists his research interests in genomics and epigenetics, and his work on the epigenetic regulation of gene expression.



Sin embargo ...

- ▶ **Errores simples** como cambios de etiquetas sensible/resistente, desplazamiento de variables ... **nos pueden pasar** a todos.
- ▶ Quizás este tipo de errores no sean tan raros

The screenshot shows a journal article from 'Genome Biology'. At the top, there's a navigation bar with links for 'HOME', 'ABOUT', 'ARTICLES' (which is underlined), and 'SUBMISSION GUIDELINES'. Below the navigation bar, there's a decorative horizontal bar composed of blue and grey segments. Underneath this, the main content starts with a red 'OPEN ACCESS' button. The title of the article is 'Gene name errors are widespread in the scientific literature'. Below the title, the authors are listed as 'Mark Ziemann, Yotam Eren and Assam El-Osta'. The journal information includes 'Genome Biology 2016 17:177 | DOI: 10.1186/s13059-016-1044-7 | © The Author(s) 2016' and the publication date 'Published: 23 August 2016'. The abstract section begins with a short summary of the findings.

- ▶ Además, análisis **big data**/ómicos conllevan **pérdida de intuición** en el análisis. Necesidad de **validación** externa de resultados.
- ▶ ¿Podemos afirmar que cualquier **publicación** está **libre de errores** de bulto sin un escrutinio externo de sus análisis?

Obligados a dudar de la literatura

- ▶ La criba de la **publicación no es garantía** de nada. Tenemos la obligación de dudar de la literatura científica.
- ▶ Artículos científicos **publicitan** investigaciones científicas y, como de todo anuncio publicitario, **hemos de dudar**.
- ▶ **Autores somos sospechosos en potencia** ya que tenemos un claro **conflicto de interés**, nuestro interés por publicar.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.



La transición que queremos conseguir:

- ▶ Como autores, nos hemos de ver obligados a pasar de trabajos **publicados** a trabajos (publicados) **reproducibles**.

CITY OF STARS LYRICS

[Sebastian's Verse: Ryan Gosling & Emma Stone]

City of stars

Are you shining just for me?

City of stars

There's so much that I can't see

Who knows?

I felt it from the first embrace I shared with you

That now our dreams

They've finally come true

[Mia's Verse: Emma Stone & Ryan Gosling]

City of stars

Just one thing everybody wants

There in the bars

And through the smokescreen of the crowded restaurants

It's love

Yes, all we're looking for is love from someone else

A rush

A glance

A kiss

CITY OF STARS

from LA LA LAND

Music by JUSTIN HURWITZ
Lyrics by BENJ PASEK
& JUSTIN PAUL

The musical score for 'CITY OF STARS' is presented in a multi-page piano-vocal score. The top page features a title card with the song name and source. The music begins with a piano introduction in a moderate tempo, followed by a vocal entry. The lyrics are written in a conversational style, reflecting the dialogue between Sebastian and Mia. The score includes dynamic markings like 'Moderato cantabile' and 'Gtr. (opt.)'. The vocal part includes several melodic phrases and harmonic progressions. The piano part provides harmonic support and includes some rhythmic patterns. The score is set against a white background with black musical notation.

- ▶ No podemos pretender **comunicar estadística sólo con palabras**, necesitamos **herramientas** más allá de la publicación de artículos (Rmarkdown, github . . .).

Estudio UKCTOCS: Reflexiones



¿Hasta qué punto la **confidencialidad** puede ser impedimento para la **reproducibilidad**?

¿Hay un **exceso de celo** en la preservación de la confidencialidad? o ¿es la confidencialidad **pretexto** para no difundir datos y/o procesos?

¿Debería exigirse **hacerse públicos** datos de estudios financiados con **fondos públicos**?

¿Cuál es el **equilibrio adecuado** entre confidencialidad y transparencia/reproducibilidad? El entente actual que preserva por defecto, y de forma desproporcionada, la confidencialidad puede no ser óptimo.

Crisis de reproducibilidad:

- ▶ En opinión de los científicos, existe incluso una **crisis de reproducibilidad** en el ámbito científico, 90% de **encuestados por Nature** así lo cree.
-



Neil deGrasse Tyson

@neiltyson

Seguir

To be scientifically literate is to empower
yourself to know when someone else is full of
shit.

[http://twitter.com/...](http://twitter.com/)

- ▶ Ser **conscientes** de la crisis de reproducibilidad es el **primer paso** para superarla !!En ellos estamos!!