

# Defnyddio tystiolaeth ymchwil

Canllaw ymarfer

# Using research evidence

A practice guide

**nesta**



Y Gynghair  
ar gyfer Tystiolaeth  
Ddefnyddiol



Alliance  
for Useful  
Evidence

## Cydnabyddiaeth

Ysgrifennwyd y canllaw hwn gan **Jonathan Breckon**, fe'i golygwyd gan **Isobel Roberts** a'i gynhyrchu gan Dîm Sgiliau Arloesi Nesta.

Diolch i'r canlynol am eu sylwadau, eu hawgrymiadau a'u cymorth gyda'r canllaw hwn: **Helen Anderson** (Proffesiwn Polisi, Dysgu'r Gwasanaeth Sifil); **Jonathan Bamber** (Nesta); **Albert Bravo-Biosca** (Nesta); Yr Athro **Paul Cairney** (University of Stirling); **Michael O'Donnell** (BOND); **Triin Edovald** (Nesta); **Caroline Kenny** (Swyddfa'r Senedd ar gyfer Gwyddoniaeth a Thechnoleg); **Geoff Mulgan** (Nesta); **Tony Munton** (Centre for Evidence Based Management ac RTK Ltd); **Nicky Miller** (College of Policing); **Helen Mthiyane** (Y Gynghrair ar gyfer Tystiolaeth Ddefnyddiol); **Jonathan Sharples** (Sefydliad Gwaddol Addysgol); **Louise Shaxson** (Sefydliad Datblygu Tramor); **Howard White** (Campbell Collaboration).

Hoffem hefyd ddiolch i'r holl gyfranogwyr a phartneriaid yn ein Gweithdai Tystiolaeth yn ystod 2014-15, gan gynnwys Dysgu'r Gwasanaeth Sifil, ACEVO (Cymdeithas Prif Weithredwyr Sefydliadau Gwirfoddol), Cyngor yr Alban ar gyfer Sefydliadau Gwirfoddol, SOLACE a Phroffesiwn Polisi Llywodraeth Cymru.

Fodd bynnag, nid yw'r safbwyntiau o reidrwydd yn adlewyrchu'r awgrymiadau a wnaed gan yr unigolion hyn na'u sefydliadau cysylltiedig. Mae unrhyw wallau yn yr adroddiad yn parhau i fod yn wallau yr awdur ei hun.

---

## Ynglŷn â'r Gynghrair ar gyfer Tystiolaeth Ddefnyddiol

Mae'r Gynghrair yn hyrwyddo'r defnydd o dystiolaeth mewn polisi ac ymarfer cymdeithasol. Rydym yn rhwydwaith agored o ragor na 3,000 o unigolion o lywodraeth, prifysgolion, elusennau, busnesau ac awdurdodau lleol yn y DU ac yn rhyngwladol.

I gofrestru ewch i: [www.alliance4usefulevidence.org/join](http://www.alliance4usefulevidence.org/join)

Cysylltwch â Chyfarwyddwr y Gynghrair ar gyfer Tystiolaeth Ddefnyddiol ar [jonathan.breckon@nesta.org.uk](mailto:jonathan.breckon@nesta.org.uk)

## Canllawiau Ymarfer Nesta

Mae'r canllaw hwn yn rhan o gyfres o Ganllawiau Ymarfer a ddatblygwyd gan dîm Sgiliau Arloesi Nesta. Cynlluniwyd y canllawiau i'ch helpu i ddysgu am ddulliau arloesi a dulliau gweithredu a'u rhoi ar waith yn eich gwaith eich hun.

I gael gwybodaeth bellach cysylltwch â: [skills@nesta.org.uk](mailto:skills@nesta.org.uk)

## Ynglŷn â Nesta

Mae Nesta yn elusen arloesi fyd-eang. Ein cenhadaeth yw tanio a thyfu syniadau newydd i wella sut mae'r byd yn gweithio i bawb. Defnyddiwn ein gwybodaeth, rhwydweithiau, cyllid a sgiliau i fynd i'r afael â heriau mawr, gan weithio mewn partneriaeth ag eraill i sicrhau bod newid yn digwydd.

Rydym yn elusen yn y DU ac mae ein gwaith yn cael ei alluogi gan waddol ariannol. Mae Nesta yn elusen gofrestredig yng Nghymru a Lloegr 1144091 ac yn yr Alban SCO42833.





## Acknowledgements

This guide was written by **Jonathan Breckon**, edited by **Isobel Roberts** and produced by Nesta's Innovation Skills team.

Thank you to the following for their comments, suggestions and help with this guide: **Helen Anderson** (Policy Profession, Civil Service Learning); **Jonathan Bamber** (Nesta); **Albert Bravo-Biosca** (Nesta); **Professor Paul Cairney** (University of Stirling); **Michael O'Donnell** (BOND); **Triin Edovald** (Nesta); **Caroline Kenny** (Parliamentary Office for Science and Technology); **Geoff Mulgan** (Nesta); **Tony Munton** (Centre for Evidence Based Management and RTK Ltd); **Nicky Miller** (College of Policing); **Helen Mthiyane** (Alliance for Useful Evidence); **Jonathan Sharples** (Education Endowment Foundation); **Louise Shaxson** (Overseas Development Institute); **Howard White** (Campbell Collaboration).

We would also like to thank all the participants and partners in our Evidence Masterclasses during 2014-15, including Civil Service Learning, ACEVO (Association of Chief Executives of Voluntary Organisations), Scottish Council for Voluntary Organisations, SOLACE and the Welsh Government Policy Profession.

The views do not, however, necessarily reflect the suggestions made by these individuals or their affiliated organisations. Any errors in the report remain the author's own.

---

## About the Alliance for Useful Evidence

The Alliance champions the use of evidence in social policy and practice. We are a free, open-access network of over 3,000 individuals from across government, universities, charities, business and local authorities in the UK and internationally.

To sign up visit: [www.alliance4usefulevidence.org/join](http://www.alliance4usefulevidence.org/join)

Contact: Director of Alliance for Useful Evidence, Jonathan Breckon on [jonathan.breckon@nesta.org.uk](mailto:jonathan.breckon@nesta.org.uk)

## Nesta's practice guides

This guide is part of a series of practice guides developed by Nesta's Innovation Skills team. The guides have been designed to help you learn about innovation methods and approaches and put them into practice in your own work.

For further information contact: [skills@nesta.org.uk](mailto:skills@nesta.org.uk)

## About Nesta

Nesta is a global innovation foundation. Our mission is to spark and grow new ideas to improve how the world works for everyone. We use our knowledge, networks, funding and skills to take on big challenges, working in partnership with others to make change happen.

We are a UK charity and our work is enabled by a financial endowment. Nesta is a registered charity in England and Wales 1144091 and Scotland SC042833.



# Defnyddio tystiolaeth ymchwil

## Canllaw ymarfer

Cyflwyniad 6

Adran A: Beth yw gwneud penderfyniadau yn seiliedig ar dystiolaeth, a pham canolbwyntio ar ymchwil? 8

Adran B: Pryd gall tystiolaeth eich helpu? 15

Adran C: 'Llunio'r gwadn fel bo'r droed' – Pa dystiolaeth ddylech chi ei dewis? 20

RHAN 1: Gwahanol fathau o ddulliau ymchwil, dyluniadau ac ymagweddau 22

RHAN 2: Sut mae barnu ansawdd ymchwil? 31

Adran D: Ble dylech chi edrych am dystiolaeth? 41

Adran E: Sut dylech chi gyfleu eich canfyddiadau? 50

Crynodeb 53

Atodiad 1: Sail resymegol i dystiolaeth a mathau o dystiolaeth sy'n ofynnol ar gyfer llunwyr polisi 55

Atodiad 2: Dyluniadau arbrofol ar gyfer gwerthuso ymyriadau cymhleth 56

Atodiad 3: Canllawiau ar ansawdd tystiolaeth, dyluniadau a dulliau 57

Ôl-nodiadau 58

# Using research evidence

## A practice guide

**Introduction** 63

**Section A: What is evidence-informed decision-making, and why focus on research?** 65

**Section B: When can evidence help you?** 72

**Section C: 'Horses for courses' – What evidence should you choose?** 77

PART 1: Different types of research methods, designs and approaches 79

PART 2: How do you judge the quality of research? 88

**Section D: Where should you look for evidence?** 97

**Section E: How should you communicate your findings?** 105

**Summary** 108

**Appendix 1: Rationales for evidence and types of evidence required for policymakers** 109

**Appendix 2: Experimental designs for evaluating complex interventions** 110

**Appendix 3: Guidance on evidence quality, designs and methods** 111

**Endnotes** 112

---

# Cyflwyniad

*"Nid oes unrhyw beth yn waeth gan y llywodraeth na bod yn hyddysg; mae'n gwneud y broses o wneud penderfyniadau yn llawer mwy cymhleth ac anodd."*

John Maynard Keynes

Gall tystiolaeth ymchwil eich helpu i ddeall beth sy'n gweithio, ym mhle, pam ac i bwy. Gall hefyd ddweud wrthyhych chi beth sydd ddim yn gweithio, a gallwch osgoi ailadrodd methiannau eraill drwy ddysgu o werthusiadau rhaglenni aflwyddiannus.

Mae tystiolaeth hefyd yn herio'r hyn rydyn ni'n ei feddwl sy'n synnwyr cyffredin. Er enghraifft, efallai eich bod yn meddwl ei fod yn syniad da cynyddu nifer yr heddlu sydd ar y strydoedd er mwyn lleihau troseddau neu lleihau maint dosbarthiadau - ond nid yw'r dystiolaeth o reidrwydd yn cefnogi hyn. Efallai y bydd mwy o heddlu yn eu lifrai yn cerdded y strydoedd yn gwneud i bobl deimlo'n fwy diogel, ond mewn gwirionedd gall dynnu'r heddlu i ffwrdd o ddatrys troseddau.<sup>1</sup> Er gwaethaf hyn, roedd y rhan fwyaf o daflenni gwleidyddol a maniffestos yn Etholiad Cyffredinol y DU yn 2015 yn honni y byddai cynyddu nifer yr heddlu sydd ar y stryd yn lleihau troseddau. Fe wnaeth y gwleidyddion anwybyddu'r dystiolaeth.<sup>2</sup>

Pan fydd cyllidebau'n dynn, ni allwn fforddio gwastraffu arian ar bolisiâu a rhaglenni sydd ddim yn gweithio. Efallai bod cael gliniaduron a gismos technegol ym mhob ysgol yn ymrwymiad fyddai'n hawlio'r penawdau, ond ychydig iawn mae'n ei wneud mewn gwirionedd er budd addysg pobl ifanc. Mewn gwirionedd bu i ymchwil gan yr OECD ganfod bod defnyddio cyfrifiaduron yn rheolaidd mewn ysgolion yn aml yn gysylltiedig â chanlyniadau academaidd is.<sup>3</sup>

Wrth gwrs bydd taflu arian at y rhan fwyaf o broblemau fel arfer yn gwneud rhyw fath o ddaioni. Ond a allai'r arian hwnnw fod wedi cael ei wario yn fwy effeithiol ar rywbeth rhatach? Mae'n ymddangos yn amlwg y byddai lleihau maint ystafelloedd dosbarth yn gwella ansawdd y dysgu a'r addysgu drwy roi hwb i'r sylw un i un mae dysgwyr yn ei gael. Ond yn gyffredinol ychydig o fudd mae'r dystiolaeth yn ei ddangos. Er mwyn cael effaith, byddai angen lleihau dosbarthiadau nodweddiadol i 15 o ddisgyblion - cam costus ar gyfer y rhai hynny sy'n rheoli cyllidebau ysgolion. Yn hytrach, mae dewisiadau eraill sy'n rhatach ac yn cael mwy o effaith, fel rhoi adborth i athrawon neu diwtora cymheiriaid.<sup>4</sup>

P'un a yw mewn gorsaf heddlu, ystafell ddosbarth ysgol neu ystafell bwrdd elusen, gall tystiolaeth eich helpu i wneud penderfyniadau gwell. Mae'n ddefnyddiol nid yn unig i ddarparu gwasanaethau rheng flaen, ond hefyd i greu sefydliadau mwy clyfar - elusennau, awdurdodau lleol, adrannau'r Llywodraeth - ac wrth ddatblygu polisiâu cenedlaethol neu ymgyrchoedd elusennol.

Rydym wedi creu'r canllaw hwn i'ch rhoi ar y trywydd iawn i ddod o hyd i ba dystiolaeth allai eich helpu. Dylai helpu i adeiladu eich hyder wrth 'lunio, cymhathu, crynhoi, dehongli â chyflwyno sylfaen tystiolaeth gref'<sup>5</sup> gan ddefnyddio ymchwil bresennol, yn ogystal â meddwl am sut y gallwch fynd ymlaen i werthuso eich gwaith eich hun.

## Pwy sy'n debygol o ddarllen y canllaw hwn?

Mae'r canllaw hwn wedi'i anelu at y sawl sy'n gwneud penderfyniadau mewn llywodraeth, elusennau, sefydliadau gwirfoddol, cyrff aelodaeth proffesiynol ac awdurdodau lleol sy'n gweithio ym maes polisi ac arfer cymdeithasol yn y DU. Nid yw wedi'i anelu at werthuswyr ac ymchwilwyr wedi'u hyfforddi, ond yn hytrach ei nod yw meithrin galw deallus am dystiolaeth ymchwil gan gynulleidfaeod ehangach.

## Sut i ddefnyddio'r canllaw hwn

Mae'r canllaw wedi'i rannu i'r pum adran ganlynol:

### Adran A: Beth yw gwneud penderfyniadau yn seiliedig ar dystiolaeth, a pham canolbwyntio ar ymchwil?

Mae **Adran A** yn trafod yr hyn rydyn ni'n ei olygu wrth wneud penderfyniadau yn seiliedig ar dystiolaeth, a pham fod ymchwil yn elfen hanfodol ohono.

### Adran B: Pryd all tystiolaeth eich helpu?

Mae **Adran B** yn edrych ar wahanol sefyllfaoedd lle gall defnyddio tystiolaeth eich helpu, yn ogystal â'r mathau o dystiolaeth y byddwch chi eu hangen ar wahanol gamau o ddatblygu.

### Adran C: 'Llunio'r gwadn fel bo'r droed' – Pa dystiolaeth dylech chi ei defnyddio?

Mae **Adran C** yn edrych ar fathau gwahanol o dystiolaeth ac yn edrych ar sut i ddewis yr un mwyaf priodol ar gyfer eich achos. Mae hefyd yn trafod sut i farnu ansawdd y dystiolaeth.

### Adran D: Ble dylech chi edrych am dystiolaeth?

Mae **Adran D** yn cynnig cyngor ac adnoddau i'ch helpu chi i ddod o hyd i'r dystiolaeth briodol i gefnogi eich achos.

### Adran E: Sut dylech chi gyfleu eich canfyddiadau?

Mae **Adran E** yn canolbwyntio ar sut i gyfleu eich neges pan fydd gennych chi'r dystiolaeth rydych chi ei hangen.

# Adran A

## Beth yw gwneud penderfyniadau yn seiliedig ar dystiolaeth, a pham canolbwyntio ar ymchwil?

**Mae'r adran hon yn trafod yr hyn rydyn ni'n ei olygu wrth wneud penderfyniadau yn seiliedig ar dystiolaeth, a pham fod ymchwil yn elfen hanfodol ohono.**

I ddechrau, gadewch i ni fod yn glir ynghylch yr hyn nad ydyn ni'n ei olygu. Nid ydym yn sôn am ddilyn casgliadau ymchwil anhyblyg yn slafaid. Bydd barn broffesiynol a ffynonellau eraill o wybodaeth - megis adborth gan eich rhanddeiliaid - bob amser yn bwysig. Nid yw'r canllaw hwn yn ymwneud â disodli barn broffesiynol, neu esgus bod atebion hawdd, ond â gwella ansawdd y dystiolaeth a ddefnyddir mewn polisi cymdeithasol.

Dechrau da i ddisgrifio'r hyn rydyn ni'n ei olygu fyddai benthycia diffiniad o fyd meddygaeth. Ddau ddegawd yn ôl, cynigodd David Sackett a'i gydweithwyr y diffiniad canlynol ac mae wedi sefyll prawf amser:

*"Meddygaeth sy'n seiliedig ar dystiolaeth yw'r defnydd cydwybodol, eglur a doeth o dystiolaeth bresennol orau wrth wneud penderfyniadau am ofal cleifion unigol. Mae arfer meddygaeth sy'n seiliedig ar dystiolaeth yn golygu integreiddio arbenigedd clinigol unigol gyda'r dystiolaeth glinigol allanol orau o ymchwil systematig."*<sup>6</sup>

Nid hon oedd yr ymgais gyntaf i ddiffinio Meddygaeth sy'n Seiliedig ar Dystiolaeth (EBM), ond mae wedi bod yn ddylanwadol ac mae'r un mor berthnasol i bolisi cymdeithasol ag ydyw i feddygaeth. Mae'n pwysleisio sut y gall ymchwil ategu barn broffesiynol neu ffynonellau eraill o wybodaeth.<sup>7</sup>

Fodd bynnag, dylai unrhyw fodel da o wneud penderfyniadau ofalu rhag rhoi barn broffesiynol ar bedestal gan y gallai arbenigwyr gael pethau'n gwbl anghywir weithiau. Yn ddiweddarach yn yr adran hon byddwch yn darllen am sut y gallwn fod yn 'afresymol yn ôl y disgwyl' ac - yn fwriadol neu'n anfwriadol - sut y gallwn wneud camgymeriadau mewn barn bwysig. Mewn arfer meddygol, gwelodd un dystiolaeth bod 'ffactorau gwybyddol' yn cyfrannu at 74 y cant o ddiagnosisau clinigol anghywir - gan arwain at anafiadau a marwolaeth.<sup>8</sup> Byddwn yn ystyried sut i liniaru'r camgymeriadau hyn mewn penodau diweddarach.

Mae modelau gwneud penderfyniadau eraill hefyd wedi pwysleisio pwysigrwydd cyfuno gwybodaeth am dystiolaeth gyda barn. Mae Gwasanaeth Sifil y DU<sup>9</sup> yn argymhell y dylai llunwyr polisi'r DU feddu ar sgiliau 'gwleidyddiaeth' a 'darparu' i ategu eu gallu i gasglu a defnyddio tystiolaeth. Mae hyn hefyd yn berthnasol i eraill y tu allan i lunio polisiâu, megis



staff mewn elusennau neu awdurdodau lleol. Efallai fod y wleidyddiaeth ychydig yn wahanol - yn cynnwys gwleidyddiaeth 'g' fach eich sector neu sefydliad yn hytrach na manifestos a phleidiau gwleidyddol - ond mae'n parhau i fod yn rhywbeth i'w feistrol. Mae darparu yn bwysig i unrhyw sefydliad, ond mae'r canllaw hwn yn canolbwyntio ar hybu tystiolaeth.

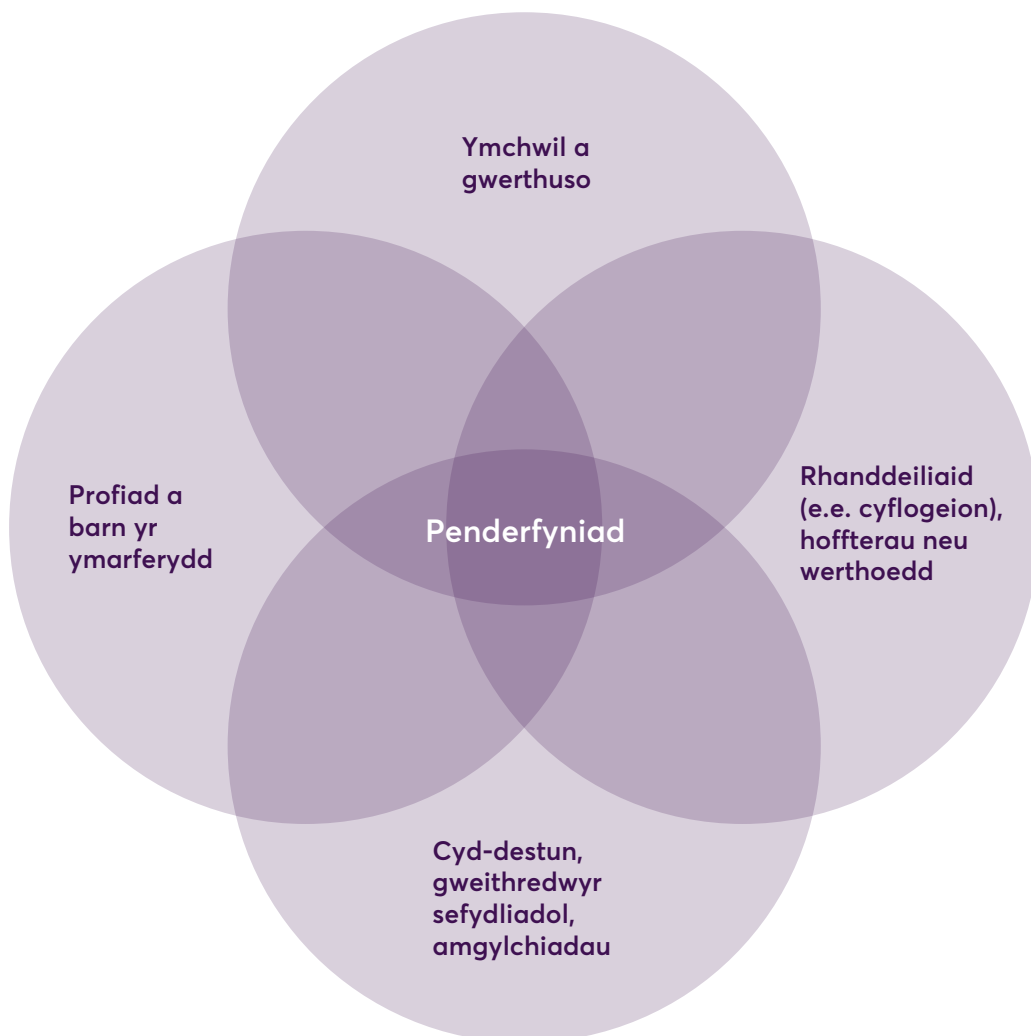
## Beth yw 'tystiolaeth', a pham rydyn ni'n canolbwyntio ar ymchwil?

Diffiniad Geiriadur Saesneg Rhydychen o 'dystiolaeth' yw: *"The available body of facts or information indicating whether a belief or proposition is true or valid."*<sup>10</sup>

Rydyn ni'n defnyddio fersiwn y geiriadur hwn oherwydd mae nifer o ddiffiniadau eraill yn dueddol o fod braidd yn ddi-fudd drwy fod yn rhy gynhwysol - gan gynnwys yr holl wybodaeth bron - neu drwy fod yn rhy haniaethol ac amwys.

Fodd bynnag, rydyn ni'n gwyrto oddi wrth y diffiniadau hyn braidd drwy ganolbwyntio ar un math o dystiolaeth. Mae Ffigur A1 yn dangos y gwahanol elfennau a ddylai fod yn rhan o'r broses o wneud penderfyniadau ar sail tystiolaeth, ond rydyn ni'n canolbwyntio ar gylch uchaf y diagram Venn: ymchwil a gwerthuso.

Ffigur A1: **Pedair elfen rheoli sy'n seiliedig ar dystiolaeth**



Mae'r rheswm dros y ffocws hwn wedi'i gyfleu'n dda gan ganllaw cynharach ar ddefnyddio tystiolaeth a gynhyrchwyd yn ystod y broses o greu UK Centre For Evidence Based Policy yr ESRC ar ddechrau'r 2000au:

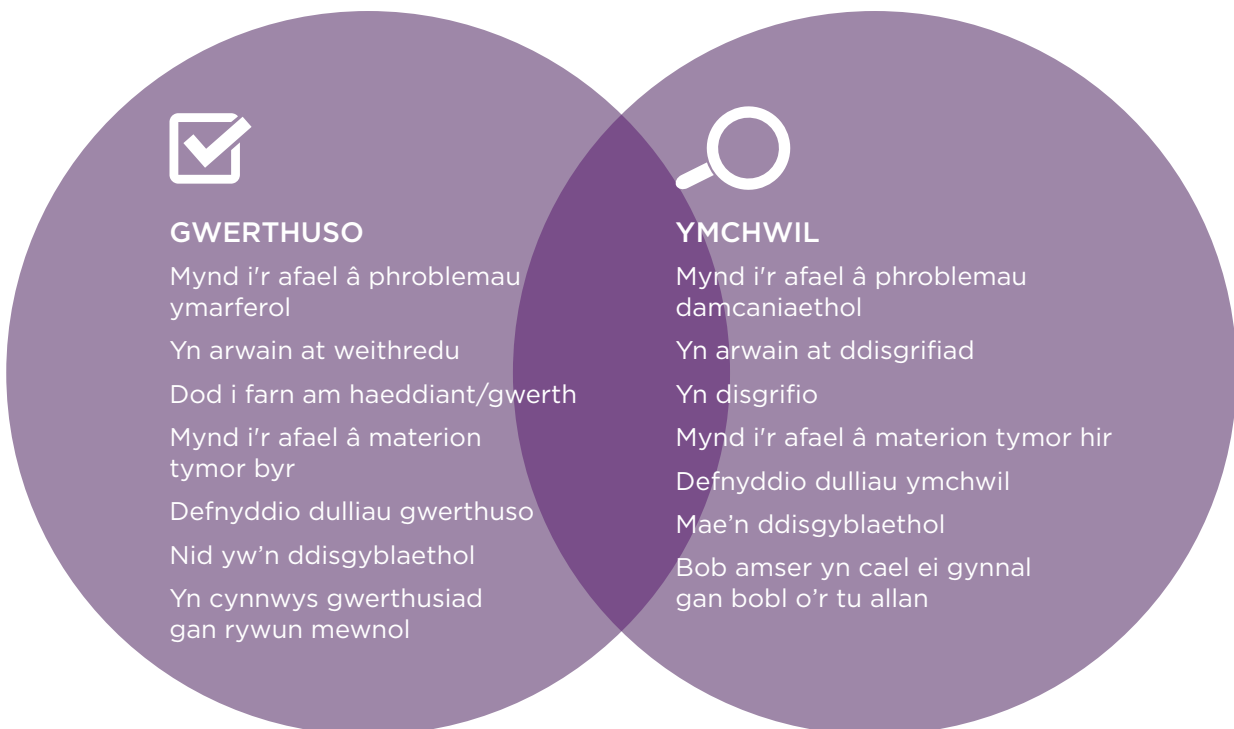
*"Wrth gyfeirio at 'dystiolaeth ymchwil', mae hyn yn cynnwys tystiolaeth o erthyglau a phapurau ymchwil a gyhoeddwyd, neu ffynonellau heb eu cyhoeddi megis gwerthusiadau a gynhaliwyd yn fewnol. Un math o dystiolaeth yw ymchwil, ond mae ganddo fanteision megis mwy o **drylwyredd**, **perthnasedd** ac **annibyniaeth** o'i gymharu â rhai mathau eraill o dystiolaeth."*<sup>11</sup>

Felly pam fod gan dystiolaeth ymchwil y manteision hyn? Fel mae awduron adroddiad *What Counts as Good Evidence?* gan y Gynghrair ar gyfer Tystiolaeth Ddefnyddiol yn ei nodi:

*"Mae cynnal a chyhoeddi ymchwil yn cynnwys dogfennu dulliau yn benodol, adolygu cymheiriaid a chraffu allanol, gan arwain at drylwyredd a bod yn agored. Mae'r nodweddion hyn yn cyfrannu at ei natur systematig ac yn helpu drwy gynnig modd o farnu didwylledd canfyddiadau. Maen nhw hefyd yn cynnig y potensial i asesu dilysrwydd un honiad o'i gymharu ag un arall."*<sup>12</sup>

Gallai mathau eraill o dystiolaeth – megis gwerthusiadau mewnlol – fod yn ddefnyddiol ac weithiau mae ganddynt y nodweddion hyn. Ond gall 'gwerthusiadau mewnlol' fod yn anodd i bobl eraill ddod o hyd iddynt; maen nhw'n aml yn gudd ar wefannau sefydliadau ac efallai nad ydynt wedi cael eu hadolygu gan gymheiriaid. Hefyd, mae gormod o werthusiadau mewnlol o ansawdd gwael (gweler tudalen 31), ac maen nhw'n fwy tebygol o fod yn destun i'r nifer o ragfarnau sy'n gallu effeithio ar ymchwil a gwerthuso (gweler tudalen 11).

Ffigur A2: Nodweddion cyffredin gwerthuso yn erbyn ymchwil<sup>13</sup>



Mae'r canllaw hwn yn canolbwyntio ar ymchwil, ond mae llawer o orgyffwrdd gyda maes gwerthuso a byddwn yn trafod rhai dulliau gwerthuso effaith yn Adran C. Rydym hefyd yn rhoi'r rhan fwyaf o sylw i ymchwil sy'n ymwneud ag effaith - p'un a yw'n rhywbeth sydd wedi cael canlyniadau cadarnhaol neu negyddol - gan fod cwestiynau ar effaith yn hanfodol i'n cynulleidfa. Maen nhw'n meddwl am ddangos eu 'heffaith' mewn elusennau, 'canlyniadau' mewn datblygu rhyngwladol neu 'beth sy'n gweithio' ar gyfer llywodraeth. Gallai'r iaith newid, ond mae'r delfryd yn aros yr un fath: i weld p'un a ydych chi wedi gwneud gwahaniaeth mewn gwirionedd.

Rydym hefyd yn amlygu ymchwil a gwerthuso parod, heb fod unrhyw angen i gynnal astudiaeth newydd sbon. Mae llawer o bobl sy'n gwneud penderfyniadau yn brin o amser ac adnoddau ac ni allant fforddio comisiynu gwaith ymchwil newydd. Yn aml mae'r angen am dystiolaeth yn awr, nid mewn blwyddyn, felly mae'r sawl sy'n gwneud penderfyniadau angen deunydd y gallant ei 'dynnu oddi ar y silff'. Yn ffodus, mae hyn yn bosibl ac rydym yn ymdrin â hyn yn Adran D.

## Arbenigwyr yn erbyn mwncïod sy'n taflu dartiau: y broblem gyda barn broffesiynol

Rheswm arall dros freintio ymchwil yw oherwydd bod barn broffesiynol yn gallu methu â chael pethau'n iawn. Bu i astudiaeth yr awdur gwleidyddol a'r seicolegydd o America, yr Athro Philip E. Tetlock,<sup>14</sup> ganfod fod yr hyn yr oedd y rhan fwyaf o arbenigwyr yn ei ddarogan yn anghywir. Fe wnaeth gasglu 80,000 o ragfynegiadau arbenigol a'u cymharu gyda'r hyn a ddigwyddodd mewn gwirionedd. Roedd y canlyniadau'n ddinistriol. Bu i academyddion, swyddogion llywodraeth, newyddiadurwyr a doethinebwyr eraill berfformio'n waeth na 'mwncïod yn taflu dartiau' wrth ddarogan y dyfodol. Yn wir, ymddangosai fod yr arbenigwyr hynny gyda gwybodaeth pwnc mwy manwl yn perfformio hyd yn oed yn waeth na'r cyfartaledd.

## Rhagfarn wybyddol

Ond nid yw arbenigwyr yn wael am ragweld y dyfodol yn unig. Gallan nhw wneud camgymeriadau yn awr hefyd. Un ffaith allweddol ym maes seicoleg yw pa mor aml mae pobl yn gwneud camgymeriad o ran barn – nid pobl leyg yn unig, ond llawfeddygon, barnwyr neu wyddonwyr ffrensig sydd wedi'u hyfforddi i lefel uchel yn gweithio mewn sefyllfaoedd byw neu farw. Rydym yn cario baich trwm o ragfarnau, rhagdybiaethau a thueddrwydd hyd yn oed.<sup>15</sup> Fel bodau dynol rydym yn afresymol yn ôl y disgwyl<sup>16</sup> a gallwn brofi hyd at 150 o ragfarnau gwybyddol sy'n amharu ar ein ffordd o feddwl. Mae rhagfarn gwybyddol yn cyfeirio at ein hanallu i fod yn gwbl wrthrychol, a gallai amlygu ei hun drwy nifer o lwybrau posibl – megis afluniad canfyddiadol, barn anghywir a dehongliadau afresymegol a/neu afresymol. Mae'n golygu y dylem fod yn wylidwrus iawn ynghylch cywirdeb penderfyniadau arbenigol.<sup>17</sup>

Hyd yn oed gyda'r bwriadau gorau, gall gweithwyr proffesiynol gael pethau'n anghywir iawn. Er enghraifft, ystyriwch un math o wall gwybyddol, tuedd i gytuno. Dyma'r duedd i weld y dystiolaeth sy'n cyd-fynd â'r hyn a gredwn, ac anwybyddu neu ddiystyru'r hyn nad yw'n cyd-fynd â'r hyn a gredwn. Gall hyd yn oed arbenigwyr profiadol syrthio i'r fagl wybyddol hon (gweler yr astudiaeth achos ar dudalen 12).

Canfu astudiaeth o weithwyr cymdeithasol yn asesu'r risg o gam-drin plant 'wrth y drws ffrynt' nad oedd eu cywirdeb yn llawer gwell na dyfalu. Tuedd i gytuno oedd un o'r ffactorau oedd ar fai am yr anhawster hwn wrth wneud penderfyniadau ar gam-drin plant yn ôl y Behavioural Insights Team a gynhaliodd yr astudiaeth ar gyfer Adran Addysg y DU.<sup>18</sup>



## Gwyddonwyr fforensig yn syrthio i fagl 'tuedd i gytuno'

Bu i'r achos o fomio trêrn yn 2004 yn ystod oriau brig Madrid ladd 191 o bobl a niweidio 1,800. Roedd awdurdodau Sbaen yn amau al-Qaeda. Yn y ras i ganfod y bomwyr, a dan gysgod 9/11, penderfynodd tri arbenigwr olion bysedd yr FBI yn hyderus fod ôl bys o fagl oedd yn cynnwys dyfeisiau bomio yn perthyn i Brandon Mayfield, cyfreithiwr o Oregon, UDA.

Treuliodd Mayfield 17 diwrnod yn nalfa'r FBI. Roedd hyn er gwaethaf pledio cyson gan awdurdodau Sbaen na allai hynny fod yn iawn. Hyd yn oed ar ôl darganfod nad oedd olion Mayfield yn cyfateb yn union i'r rhai ar y bag o fomiau, roedd arbenigwyr olion bysedd yr FBI yn parhau i resymoli'r gwahaniaethau. Dewiswyd llawer o ddarnau o dystiolaeth i gefnogi eu barn. Roedd rhai nad oeddent wir yn fforensig, megis y ffaith bod Mayfield wedi troi at grefydd Islam ar ôl priodi ei wraig o'r Aifft.

Yn dilyn y digwyddiad annifyr hwn, fe wnaeth Adran Cyfiawnder yr Unol Daleithiau orchymyn cynnal adolygiad llawn o'r achos ac yn y pen draw dywedwyd fod tuedd i gytuno wedi cyfrannu at gam-adnabod Mayfield, gan ychwanegu bod diffyg gwrthrychedd wedi arwain yr archwilwyr i weld tebygrwydd nad oedd mewn gwirionedd yn bresennol.

Dro ar ôl tro roedd yr FBI wedi addasu ffeithiau er mwyn iddynt gyd-fynd â'u damcaniaeth ei fod yn euog. Yn wir roedd cred yr FBI eu bod wedi cael y dyn cywir, er gwaethaf yr holl dystiolaeth i'r gwrthwyneb, mor gryf nes iddynt roi datganiadau camarweiniol ar lw gerbron barnwr.<sup>19</sup>

Mae bod yn ymwybodol o sut y gallwn neidio i gasgliadau yn bwysig i sicrhau ein bod yn wylidwrus o arbenigwyr. Ond mae tuedd i gytuno hefyd yn amlygu sut y gall unrhyw un – nid arbenigwyr yn unig – ddewis a dethol wrth ddefnyddio tystiolaeth ymchwil. Rydym yn chwilio am y dystiolaeth sy'n cyd-fynd â'n credoau, boed hynny'n fwriadol neu beidio. Weithiau cyfeirir at hyn fel 'tystiolaeth sy'n seiliedig ar bolisïau': dewis a dethol ac ôl-osod y dystiolaeth i'r casgliadau rydym wedi eu pennu eisoes yn ein meddyliau. Os ydym yn hael gall hyn fod yn gamgymeriad anymwybodol, ac ar ei waethaf yn gynllun bwriadol i gefnogi ein rhagfarn gan ddod o hyd i dystiolaeth sy'n cyd-fynd a chael gwared ar yr hyn nad yw'n cyd-fynd. Mae'r seicolegydd sydd wedi ennill gwobr Nobel, Daniel Kahneman, ynghyd â'i gydweithwyr Dan Lovallo ac Olivier Sibony, yn nodi 12 ffordd i fynd i'r afael â thuedd i gytuno mewn erthygl ddefnyddiol yn yr Harvard Business Review.<sup>20</sup>

Ffenomen debyg ond gwahanol yw tuedd optimistiaeth. Mae hyn yn berygl go iawn i unrhyw un sy'n sefydlu prosiect newydd mawr yn y sector cyhoeddus neu mewn elusennau – yn enwedig un sy'n cynnwys adeiladau newydd, peirianeg sifil, TG a chontractio gwaith yn allanol. Bu i un astudiaeth ddylanwadol o 258 o brosiectau seilwaith trafndiaeth gwerth US\$90 biliwn ganfod gydag arwyddocâd ystadegol llethol bod yr amcangyfrif o gost wrth gynllunio prosiectau yn hynod ac yn systematig o gamarweiniol.<sup>21</sup> Yn aml, rydym yn canolbwyntio'n ormodol ar y positif, a allai arwain at orwario mawr.

## Rhagfarnau gwybyddol perthnasol i arweinwyr polisi ac arfer cymdeithasol



### Tuedd optimistiaeth / Twyllresymeg cynllunio

Yn fwy tebygol o ddisgwyl canlyniadau cadarnhaol drwy dwyll wrth gynllunio ar gyfer y dyfodol.



### Tuedd i gytuno

Tueddiad i ddehongli a chwilio am wybodaeth sy'n cyd-fynd â chredoau rhywun ymlaen llaw.



### Tuedd ôl-ddoethineb

Tueddiad o weld digwyddiadau'r gorffennol yn fwy rhagweladwy nag yr oedden nhw cyn iddynt ddigwydd.



### Osgoi colli

Tueddiad i ffafrio osgoi colledion na chaffael enillion.



### Effaith fframio

Dod i gasgliadau gwahanol o'r un wybodaeth yn union sydd wedi'i chyflwyno mewn ffyrdd gwahanol (e.e. fyddai'n well gennych chi brydau parod sy'n '85 y cant yn ddi-fraster' neu'n '15 y cant o fraster'?)



### Hewristig argaeledd

Pan fydd pobl yn cymharu maint, amllder neu debygolrwydd problem gyda pha mor hawdd yw ei chofio neu ddychmygu.



### Hewristig cynrychioldeb

Pan fydd pobl yn goramcangyfrif tebygolrwydd digwyddiadau cofiadwy.



### Yr angen am gydlyniant

Yr awydd i sefydlu patrymau a pherthnasoedd achosol pan nad ydyn nhw o bosibl yn bodoli.<sup>22</sup>



### Rhagfarn feta-wybyddol

Y gred nad oes gennym ragfarnau!

(gyda diolch i'r Athro Rob Briner, Prifysgol Caerfaddon, a'r Athro Paul Cairney,<sup>23</sup> Prifysgol Stirling)



Nid yw hyn yn golygu bod barn broffesiynol bob amser yn anghywir. Mae ymchwilwyr eraill fel Gary Klein wedi canu clodydd barn reddfodl arbenigol yn ei waith ar 'wneud penderfyniadau naturiolaid'. Gallai barn broffesiynol a greddfau fod yn hynod werthfawr. Ond mae'n rhaid i ni fod yn ymwybodol o'u hanfanteision. Fel y gwnaeth Daniel Kahneman ei bwysleisio mewn erthygl ar y cyd â Gary Klein yn *American Psychologist*, gwelwyd yn eu hymchwil ddylanwadol bod greddf broffesiynol weithiau'n wych, ac weithiau'n ddiffygiol.<sup>24</sup>



## Trasiedi anwybodaeth broffesiynol dros ymchwil marwolaeth yn y crud

Mae marwolaeth yn y crud yn arswyd a all fod yn hunllef i rieni newydd. Yn ffodus, mae'r achosion o farwolaeth yn y crud, neu Syndrom Marwolaeth Sydyn Babanod (SIDS) i ddefnyddio'r derminoleg feddygol, wedi gostwng. Ond un o drychinebau mwyaf marwolaeth yn y crud yw'r ffaith pe byddem wedi edrych ar y gwaith ymchwil yn hytrach na gwrando ar yr arbenigwyr, gallai bywydau llawer o fabanod fod wedi cael eu hachub yn Ewrop, yn yr Unol Daleithiau ac yn Awstralia.

Yn dilyn cyngor gweithwyr iechyd proffesiynol megis yr enwog Dr Spock, fe wnaeth cenhedlaeth gyfan o rieni osod eu babanod wyneb i lawr yn y crud, gan gredu eu bod yn gwneud y peth iawn. Ond yn ôl Dr Ruth Gilbert o Goleg Prifysgol Llundain mewn erthygl yn *British Medical Journal* yn 2008, erbyn 1970 roedd llawer o dystiolaeth o ymchwil glinigol bod rhoi babanod i gysgu ar eu boliau yn cynyddu'r risg o farwolaeth yn y crud o'i gymharu â rhoi babanod ar eu cefnau.<sup>25</sup> Cafodd ymgyrch iechyd y cyhoedd, 'Nôl i Gysgu' effaith ddramatig ar farwolaeth sydyn babanod, ond ni lansiwyd yr ymgyrch tan fis Tachwedd 1991, ac nid oedd y ffordd fwy diogel o gysgu'n cael ei hargymell tan 1995.

Mae Dr Gilbert yn credu bod y cyngor i roi babanod i gysgu ar eu boliau ers bron i hanner canrif yn groes i'r dystiolaeth sydd ar gael o 1970 ymlaen bod hyn yn debygol o fod yn niweidiol. Byddai adolygiad systematig (gweler tudalen 34) o ffactorau risg y byddai'n bosibl eu hatal ar gyfer SIDS o 1970 ymlaen wedi arwain at gydnabod y niwed o gysgu ar eu boliau yn gynharach a gallai fod wedi helpu i atal rhai o'r 10,000 o farwolaethau babanod yn y DU, a'r 50,000 yn Ewrop, yr Unol Daleithiau ac Awstralia.

### Negeseuon allweddol o Adran A

- Nid ydym o blaid dilyn ymchwil yn slafaid ar draul barn broffesiynol – neu ffynonellau eraill o wybodaeth – ond hefyd **ni ddylech chwaith roi arbenigedd proffesiynol ar bedestal**. Mae ymchwil wedi dangos y gall arbenigwyr weithiau gael pethau'n ofnadwy o anghywir, ac nid ydynt yn ddiogel rhag ystod eang o dueddiadau cymdeithasol a gwybyddol.
- Gall creu tystiolaeth newydd fod yn gostus a gallai gymryd amser, ond mae digon o **dystiolaeth o ansawdd da eisoes yn bodoli** y gellir ei gael 'oddi ar y silff'.
- Un math o dystiolaeth yn unig yw ymchwil, ond mae ganddo fanteision megis **mwy o drylwyredd, perthnasedd ac annibyniaeth** o'i gymharu â mathau eraill o dystiolaeth.

# Adran B

## Pryd gall tystiolaeth eich helpu?

Mae'r adran hon yn edrych ar wahanol sefyllfaoedd lle gall defnyddio tystiolaeth eich helpu, yn ogystal â mathau o dystiolaeth y byddwch eu hangen ar wahanol gamau o ddatblygu.

Gall tystiolaeth wneud sefydliadau'n fwy effeithiol. O ymgyrchu mwy argyhoeddiadol i ennill arian grant; ac o ddatblygu gallu bwrdd i wneud penderfyniadau i wneud yn siŵr bod rhaglenni'n sicrhau canlyniadau – gall tystiolaeth gryfhau eich gwaith. Nid oes ots p'un a ydych yn sefydliad gwirfoddol bach neu'n adran llywodraeth fawr. Beth bynnag fo'r raddfa, mae math o ymchwil fforddiadwy a allai ateb eich anghenion.

Isod ceir rhai enghreifftiau o'r adegau pan allai tystiolaeth eich helpu:

Figure B1: Rhesymau dros fod angen tystiolaeth



Mae'n syniad da dechrau drwy feddwl am amseru. Bydd gwahanol dystiolaeth yn ddefnyddiol ar adegau gwahanol. Bydd angen ichi feddwl am ymchwil briodol sy'n addas ar gyfer gwahanol gamau o fywyd y rhaglen, polisi neu arfer newydd.

Yn nyddiau cynnar menter newydd, gall ymchwil nodi heriau sy'n dod i'r amlwg a maint y broblem. Er enghraifft, gallai astudiaeth hydredol (gweler Tabl C2 ar ddulliau ymchwil gwahanol) sy'n arsylwi pobl dros amser – weithiau degawdau – dynnu sylw at iechyd cyhoeddus sy'n dirywio ymysg grŵp mawr o bobl, megis iselder a morbidrwydd sy'n gysylltiedig ag unigrwydd henaint, neu iechyd a niwed seicolegol anweithgarwch corfforol mewn plant.

## Ffigur B2: Troell arloesi Nesta



Datblygwyd y droell uchod gan Nesta i ddangos camau gwahanol y broses arloesi, a gellir ei defnyddio hefyd i blotio cynnydd dull newydd ar gyfer mater cymdeithasol.<sup>26</sup> Bydd angen mathau gwahanol o dystiolaeth ar adegau gwahanol.

### 1. Archwilio cyfleoedd a heriau

Bydd gweithio drwy ymchwil gyfredol yn eich helpu i ddeall y problemau a'r cyfleoedd sy'n berthnasol i'ch maes neu fater penodol.

### 2. Cynhyrchu syniadau

Ar ôl ichi nodi eich ffocws mae'n amser chwilio am werthusiadau o ymyriadau a pholisïau sydd wedi gweithio – neu wedi methu – yn y gorffennol. Allwch chi fenthyg syniadau llwyddiannus gan eraill ac osgoi ymyriadau sydd wedi methu?

### 3. Datblygu a phrofi

Fel mae syniadau newydd yn cael eu creu, bydd yn bryd dechrau meddwl am brofi ac arbrosi gyda gwahanol ddulliau – ac am werthuso effaith y rhain. Yn adran C, rydym yn trafod Safonau Tystiolaeth sy'n dangos y daith y dylai rhaglenni newydd fynd arni o ran darparu tystiolaeth o effaith.

#### 4. Dadlau'r achos

Bydd cael tystiolaeth o effaith eich profi wedyn yn helpu gyda dadlau'r achos i'r cyllidwyr a chefnogwyr. Bydd hyn yn eich rhoi mewn sefyllfa gryfach i berswadio eich cefnogwyr a symud ymlaen i gyflawni a gweithredu.

#### 5. Cyflawni a gweithredu

Unwaith y byddwch wedi rhoi polisi, rhaglen neu brosiect ar waith mae angen ichi feddwl am sut gallwch ddangos eich effaith. Bydd hyn yn helpu i fynd â chi o sefyllfa o rethreg a dweud 'credwch fi, mae'r prosiect hwn yn gweithio' i un sy'n seiliedig ar dystiolaeth mwy dibynadwy. Bydd hefyd yn helpu gyda'ch atebolrwydd – drwy ddangos eich bod yn gwneud gwahaniaeth, eich bod yn rhoi gwerth am arian, a bod eich gwaith yn agored i'w werthuso.

#### 6. Tyfu, graddio a lledaenu

Y greal sanctaidd ar gyfer llawer o ddulliau arloesol yw ailadrodd a thyfu fel eu bod yn newid pethau ar raddfa fwy – ac y gallant o bosibl gael eu copïo'n llwyddiannus mewn lleoliadau eraill (gweler adroddiad Nesta Making It Big: Strategies for scaling social innovations<sup>27</sup>).

#### 7. Newid systemau

Mae cyflawni newidiadau system gyfan yn hynod gymhleth, ond gellir gweld hyn yn y modd y prif ffrydiwyd ailgylchu, y mudiad hosbis neu'r gwaharddiad ysmegu mewn mannau cyhoeddus. Ffordd dda i gyrraedd y raddfa hon yw cael nifer o astudiaethau annibynnol i ddangos nad yw rhywbeth yn fyrhoedlog yn unig ond yn ffordd effeithiol sy'n gweithio mewn nifer o leoedd.

### Mabwysiadu syniadau o ranbarthau neu sectorau eraill

Hyd yn oed os ydych yn canfod tystiolaeth o lwyddiant polisi neu arfer, a fyddai mabwysiadu'r syniadau hynny'n gweithio yn eich ardal chi?

Er enghraifft, roedd yn ymddangos bod digon o ymchwil gadarnhaol i gefnogi'r Nurse Family Partnership (NFP) - rhaglen wedi'i chreu yn yr Unol Daleithiau sy'n defnyddio nyrsys wedi'u hyfforddi'n arbennig i gynnal ymweliadau cartref gyda merched yn eu harddegau sy'n famau am y tro cyntaf o deuluoedd incwm isel. Mae wedi dangos llwyddiant syfrdanol, gyda gwerth dros ddau ddegawd o dystiolaeth i ategu ei heffeithiolrwydd. Ond daw llawer o'r ymchwil hwnnw o'r Unol Daleithiau, felly a fyddai'n gweithio mewn gwledydd eraill?

Yn sicr nid yw'n ymddangos felly yn y DU. Dangosodd gwerthusiad nad oedd yn debygol bod yr NFP yn gallu cael ei drosi ac nad oedd mor effeithiol yn y DU – efallai oherwydd drwy'r GIG bod mamau Prydain eisoes yn cael mwy o gefnogaeth na mamau'r Unol Daleithiau.<sup>28</sup> Mae'r enghraifft hon yn dangos bod cyd-destun yn bwysig, ac na ddylem gopïo rhaglenni o fannau eraill, oni bai eu bod yn cael eu haddasu i amgylchiadau lleol.

Ac wrth ddylunio rhaglenni newydd gyda'r bwriad o newid y ffordd rydyn ni'n gweithredu – er enghraifft, lleihau ysmegu neu leihau ymddygiad gwrthgymdeithasol – mae bob amser yn werth edrych ar ymchwil gan y rhai ym maes seicoleg gymdeithasol ac ymchwil ymddygiadol, megis y Behavioural Insights Team. Mae amrywiaeth o dechnegau i 'wthio' pobl at gamau gweithredu dymunol ar sail ymchwil gadarn megis hapbrofion wedi'u rheoli neu ymchwil ethnograffig (gweler Adran C ar ddulliau ymchwil ac ymagweddau).

## Creu damcaniaeth newid

Yn ystod camau cynnar unrhyw ymyrraeth, mae'n bwysig disgrifio'n rhesymegol beth rydych chi'n ei wneud a pham ei fod yn bwysig mewn ffordd gydlynol, glir ac argyhoeddiadol. Yn aml cyfeirir at hyn fel Damcaniaeth Newid, a'i nod yw rhoi:

*"... esboniad clir, cryno ac argyhoeddiadol o beth rydych chi'n ei wneud, yr effaith rydych chi'n gobeithio ei gael, a sut rydych chi'n credu y byddwch chi'n cyflawni hyn. Mae'n sylfaen hanfodol i unrhyw raglen, ac yn rhagofyniad ar gyfer gwerthusiad effeithiol."<sup>29</sup>*

Mae Damcaniaeth Newid yn ffordd ddefnyddiol i fod yn fwy eglur am y dystiolaeth rydych chi'n ei defnyddio gan eraill – a bod yn gliriach o ran sut rydych chi'n mynd i gael canlyniadau. Er enghraifft, os ydych am redeg rhaglen i gael plant yn fwy egniol, pa dystiolaeth sydd allan yno ar newid ymddygiadol effeithiol? Ydy adeiladu rhagor o feysydd chwarae yn ddigon? Pa gymhellion sy'n gweithio mewn gwirionedd ar blant (a'u rhieni) i'w cael oddi ar y sofa? Gallech gasglu rhai pwyntiau defnyddiol o wyddoniaeth ymddygiadol neu o werthusiad rhaglenni gweithgarwch corfforol eraill.

Mae Damcaniaeth Newid yn eich helpu i fod yn glir am eich nodau - a sut byddwch chi'n cyflawni'r nodau hynny. Mae'n helpu i osgoi gobeithio am y gorau'n unig, yn y gobaith y bydd eich arloesedd newydd yn aros ac y bydd eich rhagdybiaeth yn gywir. Mantais arall o wneud Damcaniaeth Newid yw ei fod yn gam cyntaf wrth gynllunio gwerthusiad effeithiol gan ei fod yn ceisio nodi'ch holl ganlyniadau'n gywir, sydd wedyn angen cael eu mesur.

I gael mwy o wybodaeth ar ddatblygu polisi, yn hytrach nag adeiladu elusennau neu raglenni y tu allan i'r llywodraeth, mae Atodiad 1 yn cyflwyno seiliau rhesymegol i dystiolaeth a mathau o dystiolaeth sydd eu hangen i gyfateb i'ch cwestiynau polisi.

Y neges bwysicaf yma yw meddwl am briodoldeb. Rydych eisiau dod o hyd i'r ymchwil sy'n gweddu i'ch anghenion a'r cam datblygu rydych chi arno. Bydd y math o dystiolaeth y bydd ei angen arnoch yn ystod misoedd cyntaf rhaglen newydd yn wahanol iawn i beth fydd ei angen arnoch pan fyddwch wedi eich sefydlu. Rydym yn ail-edrych ar y mater hollbwysig hwn, priodoldeb, yn yr adran nesaf.



## Negeseuon allweddol o Adran B

- Mae **ystod eang o sefyllfaoedd lle gallai tystiolaeth eich helpu**. Nid y rhai amlwg sy'n ymwneud â chasglu canlyniadau rhaglenni, polisïau ac arferion yn unig. Efallai bod manteision eraill nad ydych chi wedi meddwl amdanynt, megis creu ymgyrch fwy argyhoeddiadol, ennill arian grant neu roi'r gorau i wneud rhywbeth nad yw'n gweithio.
- Mae angen i chi feddwl am **amseru a phriodoldeb y dystiolaeth**. Meddyliwch lle rydych chi ar y droell arloesi – a pha dystiolaeth allai fod yn addas i'ch anghenion. Er enghraifft, yn nyddiau cynnar datblygu arloesedd newydd mae'n debyg y byddwch chi eisiau osgoi gwneud gwerthusiad mawr a chostus o'ch gwaith eich hun a dysgu oddi wrth eraill yn lle hynny drwy werthusiadau blaenorol neu ymchwil gwyddonol a chymdeithasol ehangach.
- Mae camau cynnar arloesi hefyd yn amser i greu **Damcaniaeth Newid**. Mae hon yn ffordd ddefnyddiol i fod yn gliriach o ran pa dystiolaeth rydych chi'n ei defnyddio a sut rydych chi am gael canlyniadau. Bydd hefyd yn eich helpu i ddylunio gwerthusiad yn effeithiol.
- Gall rhaglenni, polisïau ac arferion sydd wedi **ennill eu plwyf greu eu harbrefion neu'u gwerthusiadau eu hunain o effaith**. Wrth i arloesedd dyfu, dylech hefyd ystyried nifer o ailadroddiadau i wirio nad oedd y llwyddiant yn fyrhoedlog, ond y gallai weithio mewn lleoedd a chyd-destunau eraill.
- **Nid oes ots beth yw maint eich sefydliad**. Bellach mae llawer o ymchwil ar gael i bawb, hyd yn oed os mai sefydliad gwirfoddol bach ydych chi.

## Adran C

# 'Llunio'r gwadn fel bo'r droed' – Pa dystiolaeth ddylech chi ei dewis?

**Mae'r adran hon yn edrych ar fathau gwahanol o dystiolaeth ac yn edrych ar sut i ddewis yr un mwyaf priodol ar gyfer eich achos chi. Mae hefyd yn trafod sut i farnu ansawdd y dystiolaeth.**

Nid yw pob tystiolaeth yn gydradd. Mae rhai yn gryfach – ac yn fwy perthnasol i'ch her chi – nag eraill.

Neges allweddol yn y canllaw ymarfer hwn yw meddwl am briodoldeb. Mae gormod o bobl yn cael trafferth drwy beidio â meddwl yn glir ynghylch pa fath o ddyluniad ymchwil, dull ymchwil neu ffordd o gasglu data sydd wir am fodloni eu hanghenion. Gallwn gael ein harwain gan ddulliau.<sup>30</sup> Mewn geiriau eraill, rydyn ni'n dewis ein hoff ddull, fel holiaduron neu hap-dreialon wedi'u rheoli, oherwydd mai dyma'r hyn rydyn ni fwyaf cyfforddus ag ef yn hytrach na bod yn fwy agored i'r amrywiaeth o ddulliau sydd ar gael – a meddwl pa un fyddai'n gweddu orau i ateb ein her.

Er enghraifft, mae ymchwil arbrofol yn fwy addas i werthuso effaith a 'beth sy'n gweithio'. Roedd rhwydwaith 'Cychwyn Cadarn' o ganolfannau plant yn y DU – sy'n cynnig cymorth a chynghor ar faterion plant a theuluoedd – yn seiliedig ar werthusiadau hap-dreialon wedi'u rheoli ac adolygiadau systematig (gweler tudalen 23 i gael mwy o wybodaeth am yr agweddau hyn) o fentrau blynyddoedd cynnar yn yr Unol Daleithiau.<sup>31</sup> Ond gall mathau eraill o ymchwil helpu i roi elfennau eraill o ddealltwriaeth. Er enghraifft, ysbrydolwyd llunwyr polisi a oedd yn sefydlu Cychwyn Cadarn gan astudiaethau hydredol – ymchwil sy'n dilyn trywydd pobl dros ddegawdau – a oedd yn dangos tueddiadau hirdymor tlodi plant.

Mae'n bosibl nad yw mathau eraill o ymchwil yn edrych gymaint ar effaith, ond gallant ddatgelu pam a sut mae pethau'n gweithio.<sup>32</sup> Er enghraifft, gwelodd Cyngor Dinas Sunderland fod defnyddio ethnograffi<sup>33</sup> – lle mae'r ymchwilydd yn archwilio o safbwynt y rhai dan sylw – wedi helpu Sunderland i ailgynllunio ei raglen cyflogaeth a lles. Llwyddodd y Cyngor i ddod yn nes at anghenion ei gwsmeriaid gan ei roi ei hun yn esgidiau dinasyddion drwy ethnograffeg. Roedd modd iddynt wedyn wneud arbedion mawr drwy gael dinasyddion yn ôl i waith parhaol (gweler yr astudiaeth achos ar y dudalen nesaf).



## Dysgu drwy wneud – ethnograffeg a beth sy'n gweithio wrth ddod o hyd i swydd

Yn ôl yn 2007, roedd mwy na chwarter y boblogaeth yn Sunderland a oedd o oedran gweithio yn economaidd anweithgar ac roedd Cyngor Dinas Sunderland yn gwario swm sylweddol ar fudd-daliadau. Ond ychydig iawn o gynnydd a gâi ei wneud, felly ceisiodd y Cyngor ddefnyddio ymagwedd newydd; aethant ati i gynnal ymchwil ethnograffig i ddatgelu'r stori go iawn sy'n gysylltiedig â thaith pobl yn ôl i waith, sydd yn aml yn anodd ac yn fiwrocraataidd. Roedd ethnograffeg yn caniatáu iddynt feddwl am y gwasanaethau o safbwynt y defnyddiwr: drwy deithio a siarad gyda phobl ar y daith honno, roedd modd iddynt ddylunio cyfres o wasanaethau a oedd yn eu cefnogi drwyddo.

Yn ei gyfnod cychwynnol, roedd rhaglen Make It Work yn cefnogi dros 1,000 o bobl, gan gynhyrchu arbedion cynnar o fwy na chwarter miliwn o bunnoedd i'r cyngor. Yn ogystal â defnyddio ethnograffeg, roedd eu dull yn arbrol, yn seiliedig ar dystiolaeth gadarn ac wedi'i brofi gan brototeipio ailadroddol, gan wirio os oedd pethau'n gweithio, a dysgu'n gyflym ac yn rhad o gamgymeriadau a llwyddiannau.

Yr hyn mae'r enghraifft hon yn ei bwysleisio yw nad ydym bob amser yn gwybod beth sy'n gweithio – mae bylchau ac ansicrwydd yn y dystiolaeth. Ond gallwn gymryd ymagwedd arbrol a phrofi wrth inni fynd yn ein blaenau, yn hytrach na meddwl yn afrealistig y byddai ateb cwbl barod yn ei le i'w roi ar waith.

Roedd llwyddiant y rhaglen yn dibynnu ar barodrwydd Cyngor Sunderland i roi cynnig ar rywbeth newydd gan ddefnyddio tystiolaeth oedd yn seiliedig ar anghenion defnyddwyr lleol, i'w ariannu'n briodol a rhoi'r lle a'r caniatâd iddo arbrol. Cyfanswm cost cynnal y rhaglen oedd £180,000. Mae economegydd ar ran y cyngor wedi amcangyfrif bod y cyngor wedi osgoi cost o £435,000 wrth i gyfranogwyr ymgymryd â gwaith parhaus, gan gyfateb i arbediad cychwynnol o tua £255,000.

Mae'r elw cynnar hwn ar fuddsoddiad yn cael ei fwrw i'r cysgod gan yr arbedion hirdymor sydd i leihau diweithdra. Mae'r Gweinidog Gwladol presennol dros Ddiwygio Lles yn yr Adran Gwaith a Phensiynau wedi amcangyfrif mai economaidd resymol yw gwario £62,000 ar gefnogi person di-waith cyffredin i ddychwelyd i'r gwaith. £5,000 yn unig yw cost gyfartalog cyfranogi yn rhaglen Make It Work.

Ffynhonnell: *Radical Efficiency; different, better, lower cost public services*, Nesta 2010

Beth bynnag fo'r dull, p'un ai dreialon neu ethnograffeg, mae angen i'r math o ymchwil fodloni anghenion yr her – mewn geiriau eraill, 'llunio'r gwadn fel bo'r droed'.<sup>34</sup>

## RHAN 1: Gwahanol fathau o ddulliau ymchwil, dyluniadau ac ymagweddu

Felly, sut ydych chi'n mynd ati i ddeall pa ddull ymchwil sy'n briodol i'ch achos chi? Os nad ydych chi'n ymchwilydd, gallai fod yn anodd iawn dod wyneb yn wyneb â jargon a rhestrau diddiwedd o agweddu.<sup>35</sup> Er enghraifft, gweler yr enghreifftiau hyn:



Addaswyd o Luff et al. (2015)

Ond ni ddylai rhestrau hir fel hyn eich dychryn gan nad oes angen i chi wybod am bob math o agwedd. Ar gyfer 'defnyddiwr' ymchwil nad yw'n arbenigwr, mae'n llawer pwysicach – a haws – deall y tybiaethau sy'n sail i'r ffyrdd hyn o wneud ymchwil.<sup>36</sup>

Mae'r rhagdybiaethau hyn wedi'u casglu gan *How To Note*<sup>37</sup> yr Adran dros Ddatblygu Rhyngwladol (DFID), ac maen nhw'r un mor berthnasol i bolisi cymdeithasol ag y maen nhw i ddatblygu rhyngwladol:

- Mae rhai cynlluniau ymchwil yn fwy addas ar gyfer **dangos** presenoldeb perthynas achosol, megis dyluniadau arbrolfol a lled-arbrofol.
- Mae eraill yn fwy priodol ar gyfer **egluro** perthynas achosol o'r fath – gweler Y Tu hwnt i Arbrolfion: ymagweddau newydd at werthuso ar dudalen 28.
- Er bod rhai dyluniadau yn fwy defnyddiol ar gyfer **disgrifio** cyd-destunau gwleidyddol, cymdeithasol ac amgylcheddol – megis astudiaethau arsylwi.

I ddeall ychydig mwy am fanteision ac anfanteision yr agweddau hyn, gweler Tabl C1.

Tabl C1: **Dyluniadau, dulliau ac ymagweddau gwahanol o dystiolaeth ymchwil – trosolwg byr**

Mathau o ymchwil a gwerthuso	Beth ydyw?	Manteision	Anfanteision
<b>Hanes personol</b>	Adroddiad gan berson gyda phrofiad uniongyrchol o fater penodol, neu sydd wedi'i effeithio gan fater penodol.	Pwerau ac ar unwaith; gallai roi dealltwriaeth fyw o ddigwyddiadau sydd wedi'u cuddio oddi wrth lawer o'r boblogaeth. <sup>38</sup>	Anodd i'w ddilysu; gallai arwain at chwyddo pa mor gyffredin ydyw; gallai straeon emosiynol person cyntaf lesteirio gwerthuso beirniadol; efallai nad yw hanesion unigol yn gynrychioladol.
<b>Adborth defnyddiwr gwasanaeth</b>	Adroddiadau naratif neu arolwg o farn defnyddwyr neu anghenion a nodwyd.	Dealltwriaeth werthfawr gan y rhai hynny sy'n ddefnyddwyr; gorfodi gweithwyr proffesiynol i ganolbwyntio ar flaenoriaethau defnyddwyr y gwasanaeth.	Mae'r gydberthynas rhwng boddhad ac effeithiolrwydd gwasanaeth yn isel; efallai nad yw'r anghenion a fynegwyd yn trosi i ddefnydd gwasanaeth gwirioneddol.
<b>Astudiaeth achos unigol</b>	Archwiliad manwl a dwys o achos unigol fel cymuned, teulu, sefydliad, digwyddiad neu unigolyn.	Hawdd ac ymarferol; gall gael ei defnyddio gan ymarferwyr a'r rhai nad ydynt yn arbenigwyr; yn dda ar gyfer ymyriadau sydd wedi digwydd eisoes; o bosibl gallant nodi canlyniadau andwyol i'r ymyriad; yn helpu i ddisgrifio arloesedd; cynhyrchu rhagdybiaethau ar gyfer astudiaethau eraill.	Yn ddrwg am gasglu achosiaeth; maint bach yn golygu ei bod yn anodd cyffredinoli i lefel genedlaethol/poblogaeth.
<b>Astudiaethau achos rheoli</b>	Yn cymharu grŵp â chyflwr iechyd, er enghraifft, gyda grŵp nad oes ganddynt y chyflwr iechyd, ac yn edrych yn ôl mewn amser i weld sut mae nodweddion y ddau grŵp yn wahanol.	Angen llai o adnoddau i'w gynnal nag astudiaethau ymyriad; defnyddiol pan nad yw hap-dreialon wedi'u rheoli yn ymarferol, e.e. astudiaethau ynghlych marwolaeth yn y crud.	Prin mewn polisi cymdeithasol (gweler gwerthusiad 'seiliedig ar achos' isod sy'n gysylltiedig ond yn fwy cyffredin mewn polisi cymdeithasol), mwy cyffredin mewn epidemioleg ac iechyd; darparu llai o dystiolaeth ar gyfer casgliad achosol na hap-dreialon wedi'u rheoli; risg uchel o duedd e.e. tuedd cofio, tuedd dethol, tuedd cyfwelydd.



<p><b>Cyfranogol</b></p>	<p>Ymagwedd lle mae barn a phrofiadau rhanddeiliaid a buddiolwyr yn cael eu hamlygu. Gellir ei ddisgrifio fel dyluniadau normadol; gwerthuso cyfranogol neu ddemocrataidd; gwerthusiad ymrymuso, dysgu drwy wneud; deialog polisi; ymchwil gweithredu cydweithredol.</p>	<p>Y buddiolwyr yw'r gorau am nodi'r damcaniaethau newid mwyaf perthnasol a chanlyniadau ystyrion; mwy o botensial i fod yn foesebol a democrataidd a deall beth mae buddiolwyr ei eisiau mewn gwirionedd; mwy priodol i ymyriadau 'dan arweiniad gwerthoedd'; mwy o gyfleoedd ar gyfer mabwysiadu rhaglen oherwydd cysylltiadau agos â'r buddiolwyr; y gallu i addasu a phersonoli ymyriad.</p>	<p>Dadl nad yw'n wrthrychol o'i hanfod, ei fod yn agored i duedd ac nad ymchwil ydyw mewn gwirionedd ond yn fwy tebyg i ideoleg a gweithrediaeth gymunedol.<sup>39</sup></p>
<p><b>Seiliedig ar ddamcaniaeth</b></p>	<p>Ffordd i werthuso sy'n edrych ar yr hyn sy'n digwydd o'i gymharu â damcaniaethau sy'n bodoli eisoes neu lwybrau achosol a nodwyd yn ystod gwerthusiad. Gall fod yn gysylltiedig â gwerthuso realydd; dadansoddiad cymharol ansoddol (QCA); dadansoddi cyfraniad; olrhain proses.</p>	<p>Cryf o ran esbonio achosion; gellir ei ddefnyddio mewn meysydd mwy blêr o bolisi cymdeithasol lle gallai nifer o achosion a chyd-destunau fod yn bwysig [<i>Noder: defnyddir y categori hun yn eang iawn fan hyn. I gael rhagor o fanylion ar y maes hun gweler Stern (2015)<sup>40</sup> a White a Phillips (2012)<sup>41</sup></i>].</p>	<p>Diffyg dulliau y cytunwyd arnynt; cyfleoedd am duedd; gwan o ran amcangyfrif swm neu raddau'r effaith; ychydig o werthusiad o'r technegau a ddefnyddiwyd e.e. o'i gymharu â'r corff mawr o lenyddiaeth sy'n beirniadu hap-dreialon wedi'u rheoli.</p>
<p><b>Astudiaeth drawstoriadol</b></p>	<p>Sampl gynrychioliadol o bobl wedi'u holi ar un pwynt mewn amser. Er bod arolygon fel holiaduron a chyfweliadau strwythuredig yn cael eu defnyddio'n aml mewn dyluniad trawstoriadol, nid y rhain yw'r unig ffyrdd e.e. gellir defnyddio dadansoddi cynnwys neu ddadansoddi ystadegau swyddogol.</p>	<p>Gellir archwilio data meintiol i ganfod patrymau cysylltiad; cymharol rad a moesegol; gellir ail-adrodd yr arolwg ar ôl cyfnod o amser, gan ddangos tueddiadau'n newid dros amser (gweler Hydredol isod).</p>	<p>Sefydlu cysylltiad ar y gorau, nid achosiaeth; prin yn bosibl priodoli unrhyw newid a fesurwyd i'r ymyriad, na deall beth fyddai wedi digwydd yn absenoldeb yr ymyriad – e.e. gallai'r newid fod wedi bod yn fater ehangach fel amodau economaidd, tywydd, ymgyrchoedd y cyfryngau – yn hytrach na'r ymyriad. Anfanteision eraill: perygl o duedd cofio, tuedd dymunolrwydd cymdeithasol, tuedd ymchwilydd (Neyman); gallai meintiau grwpiau fod yn anghyfartal; gallai ffactorau sy'n drysu fod wedi'u rhannu'n anghyfartal.</p>
<p><b>Astudiaethau carfan/hydredol</b></p>	<p>Yr un sampl o bobl yn cael eu holi dros nifer o gyfnodau mewn amser, weithiau o blentyndod i henoed.</p>	<p>Ffynhonnell orau o dystiolaeth ar y cysylltiad rhwng profiad plentyndod a chanlyniadau oedolion a gall roi cefnogaeth gref i ymyriadau cynnar penodol.</p>	<p>Yn aml daw'r data i'r amlwg yn rhy hwyr ar gyfer llunio polisi; gallai aelodau'r astudiaeth dynnu'n ôl dros amser; dull drud iawn o'i gynnal dros ddegawdau.</p>

<b>Dyluniad led-arbrofol</b>	Cynigir ymyriadau gwahanol ond heb unrhyw hap-ddyraniad i grwpiau e.e. drwy ddefnyddio poblogaethau naturiol neu gyfateb achosion.	Gall gynnig tystiolaeth resymol gref o'r berthynas rhwng yr ymyriad a'r canlyniadau mesuradwy; dull pŵerus o archwilio effaith ymyriad pan fo hapddewis yn amhosibl; gellir ei ddefnyddio ar gyfer cymunedau mawr yn ogystal â grwpiau; dim angen hapddewis o'r dechrau (ex-ante), sy'n osgoi'r materion PR a moesegol sy'n ymwneud â hapddewis.	Dim modd sicrhau bod grwpiau'n cyfateb nac atal newid dros amser, ac mae hynny'n arwain at ganfyddiadau llai dibynadwy; mae technegau cyfateb yn galw am lawer o ddata mewn grwpiau ymyrraeth a grwpiau cymharu a gallai hyn gymryd amser a bod yn ddrudd; mae angen dealltwriaeth dda o'r ffactorau sydd angen eu cyfateb – heb hyn, mae'n bosibl bod gwahaniaethau systematig rhwng y dda grŵp nad ydynt yn cael eu rheoli; mae ar y dyluniadau hyn angen gwaith dadansoddol cymhleth a gwybodaeth arbenigol.
<b>Hap-dreialon wedi'u rheoli (RCT)</b>	Mae un grŵp yn derbyn ymyriad tra bo grŵp arall yn derbyn dim neu fath arall o ymyriad, gyda siawns unfath bod y naill grŵp neu'r llall yn destun treial.	Yn aml, y canfyddiadau mwyaf cadarn a dibynadwy, sy'n rhoi hyder bod unrhyw wahaniaeth a fesurir rhwng grwpiau o ganlyniad i'r ymyriad; dylai hapddyrrannu oresgyn unrhyw wahaniaeth systematig rhwng grwpiau; mwy o hyder ym maint yr effaith, a'r berthynas rhwng yr ymyriad a'r canlyniad; dull a gydnabyddir yn rhyngwladol.	Gwael am ystyried cyd-destun e.e. amgylchiadau diwylliannol, hanesyddol ac economaidd; anodd i'w cynnal ar lefel poblogaeth genedlaethol; risg pan ddefnyddir ar lefel peilot bach nad yw'n berthnasol ar lefel y boblogaeth (er bod hyn yn risg ar gyfer yr holl ddyluniadau); gallai fod yn anodd trin newidynnau i arbrofi mewn polisi cymdeithasol e.e. dosbarth, tras neu ble rydych yn byw; gall camgymeriadau wrth hapddewis achosi i'r canlyniadau fod yn annilys; gallai fod yn anodd perswadio'r rhai sy'n gwneud penderfyniadau e.e. gwleidyddion) am fanteision y dyluniad hwn; materion gwleidyddol, moesegol a chysylltiadau cyhoeddus posibl ynghylch hapddewis (e.e. rhai grwpiau ar hap yn cael ymyriadau a allai fod yn fuddiol, ac nid eraill); gall gymryd mwy o amser i'w rheoli a'u sefydlu na lled-arbrofion.
<b>Adolygiadau systematig/meta-ddadansoddi</b>	Cyfuno canlyniadau astudiaethau cymwys, gyda'r meini prawf cymhwysedd wedi'u diffinio ymlaen llaw a'r fethodoleg wedi'i hadolygu.	Ffynhonnell orau o sicrwydd bod ymyriad yn gweithio (neu nad yw'n gweithio); meta-ddadansoddi yn cydgasglu canlyniadau ystadegol; mae i adolygiadau mawr rym ystadegol sylweddol; gellir ei efelychu gan ymchwilwyr eraill; gellir ei ddefnyddio gydag unrhyw fath o ddata.	Mae angen nifer sylweddol o astudiaethau cynradd cadarn mewn maes penodol; methodoleg heb ddatblygu cymaint ar gyfer cyfod data ansoddol a llenyddiaeth 'lwyd' [Am ymagweddau ehangach ar gyfod, gweler dulliau eraill fel cyfod naratif, a chyfod yn seiliedig ar realydd].

Addaswyd o: Quality in policy impact evaluation: understanding the effects of policy from other influences. (canllaw atodol i'r Magenta Book) Trysorlys Ei Mawrhydi/DEFRA/DECC (Llywodraeth ei Mawrhydi: 2012); The Evidence Guide; Using Research and Evaluation in Social Care and Allied Professions (2006) Barnado's/What Works for Children?/Centre for Evidence-Based Social Services; a Petticrew, M. a H. Roberts (2003), 'Evidence, hierarchies and typologies: horses for courses'. Journal of Epidemiology and Community Health.' 57: 527-529 (2003); Stern, E. (2015) 'Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.' Tabl 2, t.18.

## Ymchwil arbrofol – pam yr holl ffws?

Mae ymchwil arbrofol megis hap-dreialon wedi'u rheoli (RCT) wedi cael llawer o sylw<sup>42</sup> – gormod, yn ôl rhai. Felly pam maen nhw wedi denu cymaint o sylw?

Yn aml, mae'r rhai sy'n gweithio ym maes polisi cymdeithasol eisiau gwybod os yw ymyriad wedi cael effaith. A oes cysylltiad achosol rhwng eich rhaglen newydd a'r 'canlyniad', 'traweffaith', 'effaith' a'r 'canlyniad' ar y diwedd? Er eu bod yn anodd i'w rhoi ar waith yn ymarferol, yn aml mae gan y dyluniadau arbrofol well siawns o ddilyn trywydd yr achos a'r effaith hwn.

Yn syml, gall arbrofion brofi dau grŵp nad ydynt ar hap cyn ac ar ôl ymyriad. Ond mae'n bosibl y bydd risg bod gan grwpiau duedd neu nad ydynt yn gynrychiadol o'r grwpiau o dan ystyriaeth – er enghraifft, mae'n bosibl y bydd rhywun sy'n gofyn am ymyriad yn fwy rhagdueddol tuag at ganlyniadau cadarnhaol. Efallai fod ganddynt fwy o gymhelliant, neu'n fwy iach, yn fwy hyderus, ac felly'n fwy tebygol o wneud yn dda waeth beth fo'r 'driniaeth'.

I gael gwared ar y duedd hon, gallwn ddefnyddio proses hapddewis: mae'r grwpiau a gaiff eu cymharu yn cael eu dewis ar hap yn llwyr, er enghraifft drwy dynnu tocyn. Mae'r blwch isod yn rhoi disgrifiad mwy cyflawn o'r dull hwn, a ddaw o atodiad i Magenta Book Trysorlys Ei Mawrhydi.

### Dyluniadau ymchwil arbrofol

"Bydd dyluniad arbrofol, sydd wedi'i gynnal yn briodol, yn pennu p'un a wnaeth ymyriad achosi canlyniad. Mae dyluniadau gwerthuso o'r fath yn defnyddio dyraniad ar hap i roi unedau asesu (unigolion/grwpiau) naill ai i'r ymyriad neu i'r grŵp gwrthffeithiol (a elwir yn aml yn 'grŵp dan reolaeth' mewn dyluniad arbrofol). Drwy ddefnyddio maint sampl priodol a dyraniad priodol i'r grwpiau arbrofol neu grwpiau dan reolaeth, dyma'r ffurf gryfaf o ddylunio ar gyfer gwerthusiad effaith, gan fod y dyraniad ar hap yn lleihau'r tebygolrwydd o unrhyw wahaniaethau systematig - y rhai y gwyddom amdanynt neu'r rhai na wyddom amdanynt - rhwng y grwpiau. Felly mae'n caniatáu priodoli achos ac effaith."

*O'r Magenta Guide: Trysorlys ei Mawrhydi, DECC a DEFRA (2012) 'Quality in policy impact evaluation; understanding the effects of policy from other influences.'*

Mae cyflwyno grŵp dan reolaeth yn cael gwared ar amrywiaeth eang o dueddiadau sydd fel arfer yn cymhlethu'r broses werthuso – er enghraifft, os byddwch chi'n cyflwyno cynllun 'yn ôl i'r gwaith', sut fyddwch chi'n gwybod na fyddai'r rhai hynny sy'n cael cefnogaeth ychwanegol wedi dod o hyd i swydd beth bynnag?



### Making it work - ymchwil arbrofol ar fudd-daliadau analluogrwydd

Yn 2003, cynhaliodd yr Adran Gwaith a Phensiynau hapdreial dan reolaeth i archwilio effaith tair rhaglen newydd ar gyfer y sawl sy'n hawlio Budd-dâl Analluogrwydd. Y rhain oedd: cefnogaeth yn y gwaith, cefnogaeth yn seiliedig ar eu hanghenion iechyd unigol, neu'r ddau. Costiodd y gefnogaeth ychwanegol £1,400 ar gyfartaledd, ond ni lwyddodd yr hapdreialon dan reolaeth i ganfod unrhyw fantais dros y gefnogaeth a oedd eisoes ar gael. Yn y pen draw, fe wnaeth arbed llawer o arian i'r trethdalwr gan iddo ddarparu tystiolaeth eglur nad oedd y gefnogaeth ddud ychwanegol yn cael yr effaith a fwriadwyd.

Yn fwy diweddar roedd yr Adran Gwaith a Phensiynau yn awyddus i archwilio p'un a fyddai'n bosibl lleihau dwyster y broses gofrestru ar gyfer ceiswyr gwaith ar fudd-daliadau, heb waethygu'r canlyniadau.

Mewn treial yn cynnwys dros 60,000 o bobl, cymharwyd y broses gofrestru arferol bob pythefnos gyda nifer o rai eraill a oedd yn llai dwys o ran adnoddau (e.e. cofrestru dros y ffôn, yn llai aml). Yn y treialon hynny a oedd yn ddigon mawr i ddangos effeithiau dibynadwy, gwelwyd bod yr holl opsiynau amgen i'r sefyllfa bresennol yn lleihau'r amser a gymerai pobl i ddod o hyd i waith. O ganlyniad, er gwaethaf newidiadau eraill i'r system fudd-daliadau, mae polisi'r Adran Gwaith a Phensiynau yn parhau i olygu bod gofyn i bobl gofrestru bob pythefnos.

Ffynhonnell: Swyddfa'r Cabinet (2012) *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*.

## Materion ymarferol a moesegol

Beirniadaeth gyffredin o hapdreialon dan reolaeth yw eu bod yn anymarferol, ond yn aml, mae ffyrdd o amgylch hyn. Er enghraifft, mae'n bosibl defnyddio dyluniad 'lletem raddol' (gweler Atodiad 2 ar ffyrdd gwahanol o wneud treialon) fel ateb i rai o'r problemau moesegol ac ymarferol. Yn yr agwedd hon, mae pawb, yn y pen draw, yn cael yr ymyriad – megis dull dysgu newydd mewn ysgolion neu arloesedd lleihau troseddau.

Ond gyda threialon 'lletem raddol', mae'r drefn y mae cyfranogwyr yn derbyn yr ymyriad yn cael ei benderfynu drwy loteri. Caiff ei gyflwyno'n raddol dros amser – ond mae pwy sy'n ei gael yn cael ei ddewis ar hap. Gall hyn fod yn hynod ymarferol mewn byd o lymder. Yn aml, gan nad oes digon o adnoddau i roi rhaglen newydd ar waith yn llwyr ar unwaith, gall rhoi'r rhaglen ar waith yn raddol fod yn gost-effeithiol. Mae Atodiad 2 yn cynnwys trosolwg byr o hyn a phedwar dull ymchwil arbrofol gwahanol.

## Canllawiau defnyddiol ar ymchwil arbrofol

- *Randomised controlled trials – gold standard or fool's gold? The role of experimental methods in voluntary sector impact assessment* – cyflwyniad sylfaenol deng tudalen i hap-dreialon wedi'u rheoli i elusennau gan y Cyngor Cenedlaethol Mudiadau Gwirfoddol/y Gwasanaeth Gwerthuso Elusennau
- *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials* – canllaw defnyddiol gan Swyddfa'r Cabinet
- *Better Public Services through Experimental Government* – adroddiad Nesta ar rai o'r mythau sy'n ymwneud â moeseg a chostau hap-dreialon wedi'u rheoli.
- *Developing and evaluating complex interventions: new guidance* – canllawiau rhagorol ar amrywiaeth o hap-dreialon wedi'u rheoli a ffyrdd eraill i werthuso ymyriadau cymhleth gan y Cyngor Ymchwil Feddygol (gydag enghreifftiau o faes polisi cymdeithasol, nid meddygaeth yn unig).

## Pan nad yw dewis ar hap yn briodol

Mae'n bosibl na fydd dewis testunau eich ymchwil ar hap yn briodol. Gallai fod yn gwbl afrealistig amlygu eich 'testunau arbrol' yn fwriadol i rywbeth niweidiol fel ysmegu sigarêts neu stad o dai sy'n llawn trosedd. Gallai fod yn fwy priodol defnyddio dyluniadau nad ydynt ar hap neu 'arbrolion naturiol'.

Mae astudiaeth achos marwolaeth yn y crud (gweler tudalen 14) yn enghraifft lle na fyddai neb yn ei iawn bwyll yn cynnal arbrawf – gyda rhai babanod yn cysgu ar eu boliau ac eraill ar eu cefnau i brofi p'un a ydynt yn marw ai peidio. Ond mae modd inni gymharu grwpiau sydd wedi cael profiadau gwahanol. Defnyddiodd yr ymchwil ar farwolaethau yn y crud 'astudiaethau achos wedi'u rheoli'<sup>43</sup> (gweler eitem chwech yn Nhabl C1).

Mae enghreifftiau eraill o ymchwil sydd wedi defnyddio Hapdreialon dan Reolaeth nad ydynt yn draddodiadol ond a ddefnyddiodd grwpiau cymharol i weld p'un a oes unrhyw gysylltiad. Er enghraifft, rydym wedi gallu cymharu achosion gwahanol ar ddeddfwriaeth i osgoi manteisio ar dulliau hunanladdiad, yr effaith ar reoliadau llygredd aer mewn ardaloedd gwahanol ac effaith gwahardd ysmegu mewn mannau cyhoeddus.<sup>44</sup>

## Dyluniadau lled-arbrofol

Mae amrywiaeth eang o 'ddyluniadau lled-arbrofol', yn aml gyda theitlau lletchwith megis Dyluniadau Cyfres Amser Bylchog', Dyluniad Procsi Cyn Prawf, Dyluniad Newidynnau Dibynnol Anghyferth, Dyluniad Cyfateb Patrwm, a Dyluniad Dadleoli Pwynt Atchweliad.

Roedd yr Athro Donald Campbell, cynigydd dylanwadol hap-dreialon wedi'u rheoli a dyluniadau lledarbrofol yn y 1960au yn yr Unol Daleithiau, yn aml yn cyfeirio atynt fel arbrolion "queasy" am eu bod yn achosi i buryddion arbrolol deimlo'n sâl. Ond mae dyluniadau lled-arbrofol yn dal i fod yn ddefnyddiol pan nad yw hap-dreialon wedi'u rheoli yn ymarferol. Bu'r Gynghrair ar gyfer Tystiolaeth Ddefnyddiol yn cydweithio â'r Comisiwn Ewropeaidd, JPAL Europe a'r LSE i lunio *Testing Social Policy Innovations*<sup>45</sup> sy'n mynd i'r afael â nifer o ddyluniadau lled-arbrofol cyffredin, ac mae hwn ar gael ar ein gwefan.

## Y tu hwnt i arbrolion: ymagweddau newydd at werthuso

Mae llawer o safbwyntiau presennol ar effaith wedi symud oddi wrth ddibynnu ar arbrolion yn unig.<sup>46</sup> Gallai hap-dreialon wedi'u rheoli neu led-arbrofol weithio'n dda pan fydd ymyriad syml y gellir ei brofi. Ond anaml iawn y bydd gennym ymyriadau syml o'r fath. Er enghraifft, anaml y bydd cyrff anlywodraethol sy'n gweithio mewn cymdeithas sifil yn gweithio ar eu pen eu hunain neu'n cael y cyfle i lywio 'triniaeth' arbrolol clir ar gyfer hap-dreial wedi'i reoli. Mae gwerthuswyr yn edrych ar ffyrdd eraill o wneud pethau, dulliau sy'n cynnwys symud y tu hwnt i siarad am achosiaeth un i un – ac yn meddwl am eich 'cyfraniad'.

Ym myd cymhleth polisi cymdeithasol, mae'n annhebygol mai eich rhaglen yw'r amod angenrheidiol neu ddigonol i sicrhau llwyddiant. Mae'n debygol mai un ffactor ymysg nifer ydyw, rhan o 'becyn achosol'. Mae llwyddiant y rhaglen yn dibynnu ar beth arall sy'n digwydd neu sydd wedi digwydd o'ch amgylch. Yn y gorffennol, roedd hi'n bosibl y byddai esboniad achosol o 'beth sy'n gweithio' wedi bod yn ddigon, ond heddiw mae'n llawer mwy cyffredin i ymchwilwyr gwerthuso ofyn: 'a ydyn ni wedi gwneud gwahaniaeth?'<sup>47</sup>

Mae'r ddadl dros ymagweddau o'r fath sy'n seiliedig ar ddamcaniaeth wedi'u cyflwyno'n argyhoeddiadol mewn canllaw rhagorol a gynhyrchwyd ar gyfer y Gronfa Loteri Fawr, BOND, Comic Relief a DFID. Mae'n dangos bod esboniad, yn hytrach nag achosiaeth yn unig, yn bwysig ar gyfer unrhyw werthuswr:

*"Efallai y byddwch yn dod i gasgliad (neu resymiad achosol) o werthusiad bod ariannu ar gyfer rhaglenni addysg i ferched wedi arwain at neu wedi 'achosi' incwm teuluol uwch mewn cymuned benodol. Fodd bynnag, pan mae'n dod yn amlwg nad yw rhaglenni addysgiadol tebyg bob amser yn arwain at yr un canlyniad ym mhob man, mae pobl yn dechrau gofyn 'pam?'.*

*Mewn gwerthuso effaith, fel mewn ymchwil gwyddonol, mae esboniad yn y pen draw yn dibynnu ar ddamcaniaethau da. Mae agor y 'blwch du' sy'n cysylltu'r 'achosion' a'r 'effeithiau' yn galw am wahanol fathau o ddadansoddi, sef yr hyn a fwriedir i 'ddamcaniaethau newid' a 'damcaniaeth rhaglen' eu cefnogi.*

*Mae datblygiadau mewn gwerthuso effaith hefyd wedi achosi i werthuswyr fod yn ymwybodol bod angen iddynt ddefnyddio damcaniaethau cymuned, cymdeithasol ac economaidd ehangach er mwyn dehongli data cymhleth neu ddata sy'n aml yn ddryslyd neu'n groes hyd yn oed."*

Ffynhonnell: Stern, E (2015) *Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector*. Wedi'i baratoi ar gyfer Bond, Comic Relief, Y Gronfa Loteri Fawr

Mae amrywiaeth o ymagweddau sy'n dod yn fwy poblogaidd ymysg gwerthuswyr, megis dyluniadau sy'n seiliedig ar ddamcaniaeth<sup>48</sup> a dyluniadau sy'n seiliedig ar achosion, gwerthusiad realydd,<sup>49</sup> dadansoddi cyfraniad,<sup>50</sup> olrhain proses<sup>51</sup> neu ddadansoddi ansoddol cymharol.<sup>52</sup>

Yn hytrach na gorgyffredinoli ynglŷn â'r dulliau hyn, mae'n werth ystyried un enghraifft: dadansoddi cyfraniad. Nid yw'r math hwn o ymchwil yn ceisio profi bod un ffactor – polisi er enghraifft – wedi 'achosi' y canlyniad a ddymunir, ond yn hytrach yn archwilio'r cyfraniad mae polisi'n ei wneud i'r canlyniadau a arsylwir. Mae'n cydnabod bod effeithiau'n cael eu creu gan nifer o achosion yr un pryd, nid un yn unig. Mae gan Lywodraeth yr Alban ganllaw byr<sup>53</sup> ar y dechneg hon ac mae'n dweud y gallai fod yn ddefnyddiol pan nad yw arbrofion yn ymarferol:

*"Drwy ddatblygu 'damcaniaeth newid' sy'n dangos y cysylltiadau rhwng y gweithgareddau, y canlyniadau a chyd-destunau'r polisi a chasglu tystiolaeth o ffynonellau amrywiol i brofi'r ddamcaniaeth hon, y nod yw creu 'stori perfformiad' gredadwy. Gallai hyn ddangos p'un a oedd polisiâu a rhaglenni'n ffactor dylanwadol pwysig mewn gwirionedd wrth roi hwb i newid, efallai ynghyd â ffactorau eraill."*

Efallai nad yw'r holl ymagweddau hyn yn newydd.<sup>54</sup> Yn wir, mae cael damcaniaeth dda bob amser wedi bod wrth wraidd gwyddoniaeth dda – felly nid yw ymagweddau sy'n 'seiliedig ar ddamcaniaeth' yn ymddangos yn bethau mor newydd â hynny. Ond mae'r ffyrdd hyn o werthuso effaith wedi tyfu o ran poblogrwydd ac maen nhw'n helpu gwerthuswyr i roi sylw i fwy nag un achosiaeth.



## A oedd hi werth ei wneud? Gwerth dadansoddi cost a budd

Os ydych chi'n taflu llawer o arian at unrhyw broblem gymdeithasol, mae'n debyg y byddwch chi'n cael rhyw fudd. Ond ydych chi'n cael gwerth am arian? Pan fydd cyllidebau'n dynn, mae angen i ni wneud penderfyniadau ariannol anodd o ran p'un a yw un peth yn fwy gwerthfawr o'i gymharu ag un arall. Mae llunwyr polisi, comisiynwyr gwasanaethau, cyrff sy'n creu grantiau ac arianwyr elusennol oll yn gofyn am fwy o fesurau effaith ar gyfer pob ceiniog maen nhw'n ei dosbarthu. Gall technegau fel dadansoddi cost a budd helpu i roi dealltwriaeth o werth. Mae'n arwain at briodoli gwerth ariannol i effaith, sy'n golygu bod nifer o elusennau a chyrrff cyhoeddus yn amlwg yn chwilfrydig yn yr hinsawdd ariannu presennol.<sup>55</sup>

Mae'n bwysig nodi bod amrywiaeth o dechnegau. Un ffordd yn unig yw dadansoddi cost a budd o gymharu costau gyda buddiannau ariannol. Mae'n werth crybwyll dwy ymagwedd arall: dadansoddi cost-effeithiolrwydd a dadansoddi cost a chanlyniad. Mae'r gwahaniaethau rhwng dadansoddi cost-effeithiolrwydd a dadansoddi cost a chanlyniad yn eithaf technegol ac maen nhw'n addas i arbenigwyr, ond y gwahaniaeth cyffredinol gyda dadansoddi cost a budd yw eu bod yn mesur canlyniadau mewn unedau yn hytrach na ffigurau ariannol. Mae technegau eraill wedi cael eu datblygu'n benodol ar gyfer gwerth cymdeithasol ac amgylcheddol, megis elw cymdeithasol ar fuddsoddiad.<sup>56</sup>

Oherwydd yr ystod eang hon o ddulliau, os ydych am gynnal dadansoddiad cost a budd ar un o'ch rhaglenni gwaith eich hun, rydym yn argymhell eich bod yn defnyddio teclyn sydd eisoes yn cael ei ddefnyddio a'i gydnabod yn eang – gweler isod am ddetholiad o ganllawiau ar ddadansoddi cost a budd.

## Canllawiau ar ddadansoddi cost a budd

Mae'r *Green Book*<sup>57</sup> gan Drysorlys Ei Mawrhydi – canllaw ar gyfer arfarnu a gwerthuso polisiâu, rhaglenni a phrosiectau sy'n berthnasol i elusennau a darparwyr gwasanaeth yn ogystal â llywodraeth – yn cynnwys model dadansoddi cost a budd a ddefnyddir yn aml yn y sector cyhoeddus.

Mae angen i sefydliadau sy'n dymuno datblygu'r math hwn o ddadansoddiad gasglu data ar elfennau fel dangosyddion perfformiad allweddol, costau cyflenwi gwasanaeth, a (lle y bo'n bosibl) canlyniadau neu effaith ar ddefnyddwyr gwasanaeth. Mae canllawiau ar ba ddata i'w gasglu i'w weld mewn cyhoeddiad arall gan Lywodraeth y DU o'r enw *Choosing the Right FABRIC*,<sup>58</sup> ac mae nifer o sefydliadau sy'n gweithredu yn y sector cyhoeddus wedi defnyddio dull o gasglu a dadansoddi data perfformiad fel a nodir yn y ddogfen hon.

Canllaw defnyddiol arall, yn enwedig ar gyfer y rhai sy'n gweithio ar brosiectau lleol ac sydd â chyllidebau dadansoddi tynn, yw *Supporting public service transformation: cost benefit analysis guidance for local partnerships* a gynhyrchwyd gan Drysorlys Ei Mawrhydi, Public Service Transformation Network a New Economy in Manchester.<sup>59</sup> Mae wedi'i ddylunio i symleiddio a lleihau cost cynnal dadansoddiad cost a budd ar gyfer rhaglenni lleol sydd â'r nod o wella gwasanaethau cyhoeddus lle gallai adnoddau dadansoddi ac ymchwil fod yn gymharol gyfyngedig. Mae'n berthnasol i elusennau a busnes, yn ogystal â'r sector cyhoeddus.

Ar gyfer elusennau, bydd dogfen y Gronfa Loteri Fawr, *A Guide to Cost Benefit Analysis' Wellbeing Programme: Evaluation and Learning* hefyd yn ddefnyddiol.



## Negeseuon allweddol ar gyfer Rhan 1 Adran C

- **Nid yw pob tystiolaeth yn gyfartal.** Mae rhai o ansawdd gwell a byddant yn fwy priodol i'ch her.
- Beth bynnag fo'r dull, p'un ai'n hapdreial wedi'i reoli neu ethnograffeg, mae angen i'r math o ymchwil **fod yn addas i anghenion yr her** – hynny yw: 'llunio'r gwadn fel bo'r droed'.<sup>60</sup>
- Peidiwch â **chael eich dychryn gan restr hir o ddulliau ymchwil**, dyluniadau ac ymagweddau. Pwysicach yw deall y tybiaethau sy'n sail i'r ffyrdd hyn o wneud ymchwil.
- Mae rhai dyluniadau ymchwil yn fwy addas ar gyfer dangos presenoldeb **perthynas achosol**, megis dyluniadau arbrofol a dyluniadau lled-arbrofol.
- Mae ymagweddau eraill at ymchwil yn fwy priodol i esbonio perthnasoedd achosol o'r fath. Mae **gwerthuso ar sail damcaniaeth** a thechnegau megis dadansoddi cyfraniad yn fwyfwy poblogaidd ymysg gwerthuswyr. Gallai'r technegau hyn fod yn ddefnyddiol pan fydd hi'n anodd arbrofi neu'n amhosibl priodoli eich polisi neu raglen unigol i unrhyw ganlyniad unigol clir.
- Nid yw'n ymwneud â ph'un a yw'ch ymyriad wedi gweithio ai peidio, ond p'un a oedd yn cynnig gwerth am arian. Defnyddiwch dechnegau megis **dadansoddi cost a budd** i ddeall gwerth ariannol eich effaith.

## RHAN 2: Sut mae barnu ansawdd ymchwil?

Ffordd arall o'ch helpu i ddewis pa fath o ymchwil rydych chi ei angen yw gofyn cwestiwn gwahanol: pa ymchwil allwch chi ymddiried ynnddi? Beth sy'n dystiolaeth ddigon da i weddu i'ch anghenion? Efallai mai'r cyfnodolion academaidd gorau fel *Nature*, *BMJ* neu'r *Quarterly Journal of Economics* yw'r rhain. Ond beth am y nifer o werthusiadau mewnol a gynhelir gan elusennau, awdurdodau lleol a llywodraeth? Mae'n rhaid bod lle i'r rhain ar y bwrdd-tystiolaeth hefyd, onid oes?

Mae bod mewn cyfnodolyn ymchwil wedi'i adolygu gan gymheiriaid yn un ffordd o deimlo'n hyderus am yr ymchwil, ond nid yw'n sicrwydd. Mewn papur enwog, fe achosodd John Ioannidis o Brifysgol Stanford gynnwrf drwy ddadlau bod y rhan fwyaf o ganfyddiadau sydd wedi'u cyhoeddi fwy na thebyg yn anghywir. Archwiliodd y papurau oedd wedi'u dyfynnu fwyaf (1,000+ o ddyfyniadau) yn y cyfnodolion meddygol a ystyrir orau yn y byd – a ddeuai'n bennaf o *The Lancet*, *New England Journal of Medicine* a'r *Journal of the American Medical Association*.<sup>61</sup> O'r rhai hynny a brofwyd oedd yn honni eu bod yn effeithlon, canfuwyd bod 41 y cant naill ai'n anghywir, neu bod yr effaith yn llawer llai na'r hyn a awgrymwyd gan yr astudiaeth wreiddiol. Mae astudiaeth Ioannidis yn ddegawd oed, ond yn 2013 dywedodd wrth yr International Congress on Peer Review and Biomedical Publication nad oedd y broblem wedi mynd.<sup>62</sup>

### Ymchwil wedi'i hadolygu gan gymheiriaid: man cyfforddus y sawl sy'n gwneud penderfyniadau?

Mae'n rhaid i ni hefyd fod yn ymwybodol bod adolygu gan gymheiriaid – congllfaen cyfnodolion academaidd – yn bell o fod yn berffaith. Mae'n bosibl cael tueddiadau anymwybodol fel 'corllannu', lle mae ymddygiad adolygwyr yn cael ei ddylanwadu gan ymddygiad eu cymheiriaid. A gall y sefyllfa gyfan gael ei hystumio gan 'duedd cyhoeddiad': mae gan ganlyniadau cadarnhaol well cyfle i gael eu cyhoeddi,<sup>63</sup> tra bod data negyddol

yn cael ei roi ar y silff. Yn ôl arolwg gan Brifysgol Stanford, nad yw'r rhan fwyaf o'r 'astudiaethau nwl' ym maes cymdeithaseg yn gweld golau dydd: 20 y cant yn unig oedd wedi ymddangos mewn cyfnodolyn, ac roedd 65 y cant heb eu llunio'n bapurau hyd yn oed.<sup>64</sup>

Gallai hyn arwain at ganlyniadau difrifol ar gyfer y rhai sy'n gwneud penderfyniadau ac sy'n adolygu tystiolaeth. Os nad ydych chi'n gweld yr holl astudiaethau negyddol ar ymyriad newydd, gallech ddod i'r casgliad anghywir fod popeth yn dda. Ym maes iechyd clinigol, gallai cuddio canlyniadau treialon clinigol o gyffuriau megis Tamiflu ladd yn llythrennol.<sup>65</sup>

Er hynny, er gwaethaf y problemau hyn, i'r rhan fwyaf, adolygu gan gymheiriaid yw'r safon aur o hyd, ac yn ffordd o rwystro gwaith gwael.<sup>66</sup> Mae defnyddio cyfnodolyn a fu'n destun adolygiad cymheiriaid gan arbenigwyr eraill yn un ffordd i'ch helpu i fod rhywfaint yn fwy hyderus y gallwch ymddiried yn yr ymchwil. Gallai gymryd blynyddoedd o hyd cyn i ymchwil gael ei chyhoeddi a gall hynny fod yn rhy hir os oes angen i chi wneud penderfyniad cyflym, ond mae rhai rheolaethau ansawdd ynghlwm wrth adolygiad cymheiriaid i'ch helpu i deimlo'n hyderus.

Os nad fu'r ymchwil drwy gyfnodolyn wedi'i adolygu gan gymheiriaid, efallai bod sianeli eraill. Ar gyfer y rhai hynny sy'n gweithio mewn llywodraeth, argymhellir eich bod yn ymgynghori â chymdeithasau dysgedig (megis y Gymdeithas Frenhinol, y Gymdeithas Ystadegol Frenhinol, yr Academi Brydeinig ac eraill) neu bwyllgorau gwyddonol neu arbenigol yn bileri call ar gyfer ansawdd.<sup>67</sup> Yn y sector elusennol, mae cyrff fel **Evaluation Support Scotland** wedi cynghori defnyddio rhwydweithiau cymheiriaid a siarad gyda sefydliadau eraill o'r un feddylfryd.<sup>68</sup>

## Cadarnhau diffiniad o ymchwil 'o safon uchel'

Gallai adolygiad gan gymheiriaid roi ychydig o gysur i ni. Ond beth ydych chi am ei wneud os ydych chi am gynnwys tystiolaeth ehangach nad yw wedi'i wirio gan arbenigwyr eraill? Sut ydych chi'n penderfynu a yw'n cyrraedd y safon?

Yn amlwg, byddai'n wirion hepgor 'llynyddiaeth lwyd' bwysig gan gyrff ymchwil megis Cymdeithas Frenhinol Caeredin, Cronfa'r Brenin neu'r Sefydliad Astudiaethau Cyllid oherwydd nad oedd mewn cyfnodolyn wedi'i adolygu gan gymheiriaid. Neu hepgor y llif cyfoethog o werthusiadau llywodraeth neu elusennau nad ydynt byth yn gweld golau dydd drwy gael eu cyhoeddi gan sefydliadau academaidd. Ond mae'n teimlo'n dasg frawychus ar ôl darllen adroddiad damniol y Swyddfa Archwilio Genedlaethol ar werthuso.<sup>69</sup> O edrych ar lywodraeth ganolog, bu iddo ganfod mai llai na hanner y gwerthusiadau a archwiliwyd a oedd wedi cynnig tystiolaeth gadarn ynghylch effaith polisi – gan olygu na ellid dibynnu ar y canfyddiadau.

Dechrau da wrth geisio asesu ansawdd yw ei ddiffinio. Un o'r problemau, fodd bynnag, yw bod ymadroddion fel 'ansawdd', 'safonau', 'cadernid' a 'chryfder' yn cael eu defnyddio drwy'r amser fel pe baent yn ymglyfnewidiol, heb ddiffinio'n glir beth ydynt. Mae hyn yn arwain at lawer o flerwch a chamddealltwriaeth.

Er enghraifft, mewn rhai canllawiau,<sup>70</sup> mae 'ansawdd' yr ymchwil yn golygu defnyddio mathau penodol o ddyluniad a dull – megis hap-dreialon wedi'u rheoli. Mae'r ffocws hwn ar ansawdd yn gyfeiriad bwriadol at rai o'r ymagweddu clinigol a meddygol ffurfiol ynghylch asesu ansawdd tystiolaeth, megis GRADE neu system Scientific Maryland Scale.<sup>71</sup>

Ond i eraill, gall ansawdd olygu arfer da o ran sut rydych chi'n cofnodi<sup>72</sup> eich ymchwil, neu eich gonestrwydd<sup>73</sup> o ran peidio â ffugio eich data – fel a welwyd yn achos sgandal academaidd y seicolegydd o Ddenmarc, Diederik Stapel, a fu'n ffugio data am flynyddoedd a'i gyhoeddi mewn o leiaf 30 o bapurau wedi'u hadolygu gan gymheiriaid.<sup>74</sup> Gan achosi

hyd yn oed mwy o ddryswch, mae rhai ymchwilwyr wedi barnu bod ansawdd yn rhywbeth sy'n ymwneud â defnyddioldeb – pa mor berthnasol oedd eich gwaith i bolisi ac arfer.<sup>75</sup> Er bod defnyddioldeb yn nod clodwiw, gallai ymddangos yn ddryslud i'r rhai nad ydynt yn arbenigwyr rhoi ansawdd yn gyfystyr â defnyddioldeb.

Fodd bynnag, yn y rhan fwyaf o'r llenyddiaeth, canolbwyntia trafodaethau am ansawdd ymchwil ar ddyluniadau methodolegol. Mae dulliau penodol megis hap-dreialon wedi'u rheoli wedi ennill y safle hwn o barch fel ymchwil o 'ansawdd uchel'. Yn ôl atodiad i'r *Magenta Book*<sup>76</sup> gan Drysorlys Ei Mawrhydi – 'beibl' gwerthuso – mae dyluniadau arbrofol a dyluniadau lled-arbrofol yn dod yn agosach ati o ran gweld p'un a yw ymyriadau'n 'achosi' canlyniadau (gweler disgrifiad byr o'r dyluniadau hyn yn Atodiad 2).

## Magenta Book a Green Book Trysorlys Ei Mawrhydi

Wrth feddwl am beth yw tystiolaeth gref, mae'n amhosibl anwybyddu dwy gyfrol o ganllawiau Trysorlys Ei Mawrhydi – *Magenta Book*<sup>77</sup> a'r *Green Book*.<sup>78</sup> Er bod y ddau'n ofnadwy o hir – 141 o dudalennau a 118 o dudalennau – maen nhw'n ddarllenadwy (popeth yn gymharol) ac maen nhw'n osgoi gormod o iaith dechnegol am economeg a gwerthuso. Maen nhw'n ddarllen hanfodol i unrhyw un sy'n llunio barn ar gryfder tystiolaeth.

Er bod y ddau o werth ar gyfer asesu tystiolaeth, mae'r *Magenta Book* yn hynod ddefnyddiol wrth gynllunio eich gwerthusiadau eich hun neu adolygu tystiolaeth a grëwyd gan eraill, lle mae'r *Green Book* yn canolbwyntio mwy ar y dyfodol ac efallai'n fwy perthnasol i swyddogion mewn llywodraeth ganolog – gan gynnig cymorth ar sut i adeiladu achos busnes da ar gyfer polisi newydd, er enghraifft.

Ysgrifennwyd y *Magenta Book* hefyd gyda chynulleidfaoedd ehangach mewn golwg na Whitehall ac felly mae'n arbennig o berthnasol i awdurdodau lleol, elusennau, darparwyr gwasanaethau cyhoeddus neu ymgynghorwyr a gwerthuswyr allanol sydd eisiau gweithio gyda'r llywodraeth.



## Y perygl o ddefnyddio arbrofion 'gwan' yn sail i bolisi – achos ffoneg synthetig

Mae'r polisi cyfredol ar gyfer dysgu plant i ddarllen yn seiliedig ar ffoneg synthetig, lle mae plant yn dysgu cyfateb synau i lythrennau a grwpiau o lythrennau. Gofynnwyd i dîm a gomisiynwyd gan yr Adran Addysg a Sgiliau a dan arweiniad yr Athro Carole Torgerson, ym Mhrifysgol Efrog ar y pryd, adolygu'r dystiolaeth ar ffoneg synthetig.

Er bod rhywfaint o dystiolaeth addawol o'r Unol Daleithiau ac astudiaeth fechan yn Clackmannanshire yn yr Alban, yn dangos bod yr ymagwedd wedi gweithio, roedd y dystiolaeth yn gyfyngedig. Bu i'r adolygiad llywodraeth ganfod deuddeg treial arbrofol bach yn unig, gyda'r mwyaf o'r rheiny'n cynnwys 120 o blant.<sup>79</sup> Roeddynt yn annog pwyll wrth lunio polisi cenedlaethol. Awgrymodd Torgerson yn ei hadroddiad y dylai'r llywodraeth ei roi ar waith yn raddol, gan hapddewis yr ardaloedd cyntaf i fanteisio mewn treial.

Ond, anwybyddwyd y cyngor hwn yn ôl Torgerson, a daeth yn bolisi ar unwaith: "O ganlyniad, rydyn ni'n parhau i fod yn y niwl o ran p'un a yw ffoneg yn gweithio ai peidio fel y brif strategaeth addysgu i bob plentyn," dywedodd. "Mae rhywfaint o'r gwaith gwerthuso diweddar wedi dangos ei bod yn bosibl nad yw ffoneg synthetig yn cael yr effaith a obeithiwyd. Pe byddem wedi cynnal hap-dreial, byddem wedi gwybod cyn i'r polisi fod yn genedlaethol."<sup>80</sup>

Mae'r *Magenta Book* yn nodi bod cysylltiad uniongyrchol rhwng ansawdd a dyluniad ymchwil wrth feddwl am gwestiynau achosol.<sup>81</sup> Mae gwerthusiad effaith o ansawdd uchel yn ymwneud â mater priodoli – h.y. a oes modd priodoli mai eich ymyriad chi yw achos y canlyniadau? – ac i wneud hyn mae arnoch angen grŵp cymharu neu grŵp rheoli. Wedyn, gallwch fod yn fwy hyderus bod eich syniad da penodol wedi gwella gwasanaethau plant, wedi lleihau trosedd, neu wedi helpu ceiswyr gwaith i ddychwelyd i waith, er enghraifft.

Dylem nodi yn y fan hon bod dadleuon helaeth a thanllyd am yr hyn mae achosiaeth a phriodoli'n ei olygu.<sup>82</sup> Byddai llawer o ymchwilwyr yn dadlau'n erbyn rhoi safle breintiedig i ddyluniadau arbrol am eu bod yn anymarferol ac am nad ydynt bob amser yn dweud wrthyhch pam nad oes modd i rywbeth sy'n gweithio mewn un lle weithio mewn lleoedd eraill.<sup>83</sup>

Efallai nad ymchwilwyr arbenigol yn unig sy'n dadlau'n erbyn y diffiniad hwn o ansawdd. I unrhyw un y tu allan i faes gwerthuso, gallai ymddangos yn rhyfedd dweud bod ansawdd yn gydradd â dyluniadau arbrol. Mae'n gwbl bosibl cael ymchwil o ansawdd uchel mewn ymagweddau eraill, fel ethnograffeg, grwpiau ffocws neu arolygon ar-lein. Y mae hefyd yr un mor hawdd cael hap-dreialon wedi'u rheoli o ansawdd isel ag mewn unrhyw fath arall o ymchwil.

## Ymchwil 'ansoddol' o ansawdd uchel

Mae nifer wedi canu clodydd ymchwil 'ansoddol'<sup>84</sup> o ansawdd uchel ar gyfer llunwyr polisi (gweler tudalen 19 ar ddefnydd Cyngor Dinas Sunderland ar ethnograffeg). Dywed yr arbenigwr gwerthuso profiadol yn NatCen, sefydliad ymchwil cymdeithasol annibynnol, mai 'gwyddor gymdeithasol, drylwyr' yw'r dystiolaeth fwyaf perswadiol. Mae'n dadlau na ddylem freintio tystiolaeth feintiol dros dystiolaeth ansoddol, nac i'r gwrthwyneb chwaith. Gall tystiolaeth ansoddol fod yr un mor wyddonol gredadwy â thystiolaeth feintiol.<sup>85</sup> Ond yr hyn y dylem ei wneud yw:

*"...breintio tystiolaeth sy'n cynnwys casglu a dadansoddi data, drwy ddefnyddio methodoleg systematig, ddiffiniedig y mae modd ei hailadrodd, gan gynhyrchu data clir, profi cwestiynau gwerthuso sydd wedi'u diffinio ymlaen llaw, a sicrhau bod canlyniadau ar gael yn gyhoeddus."*

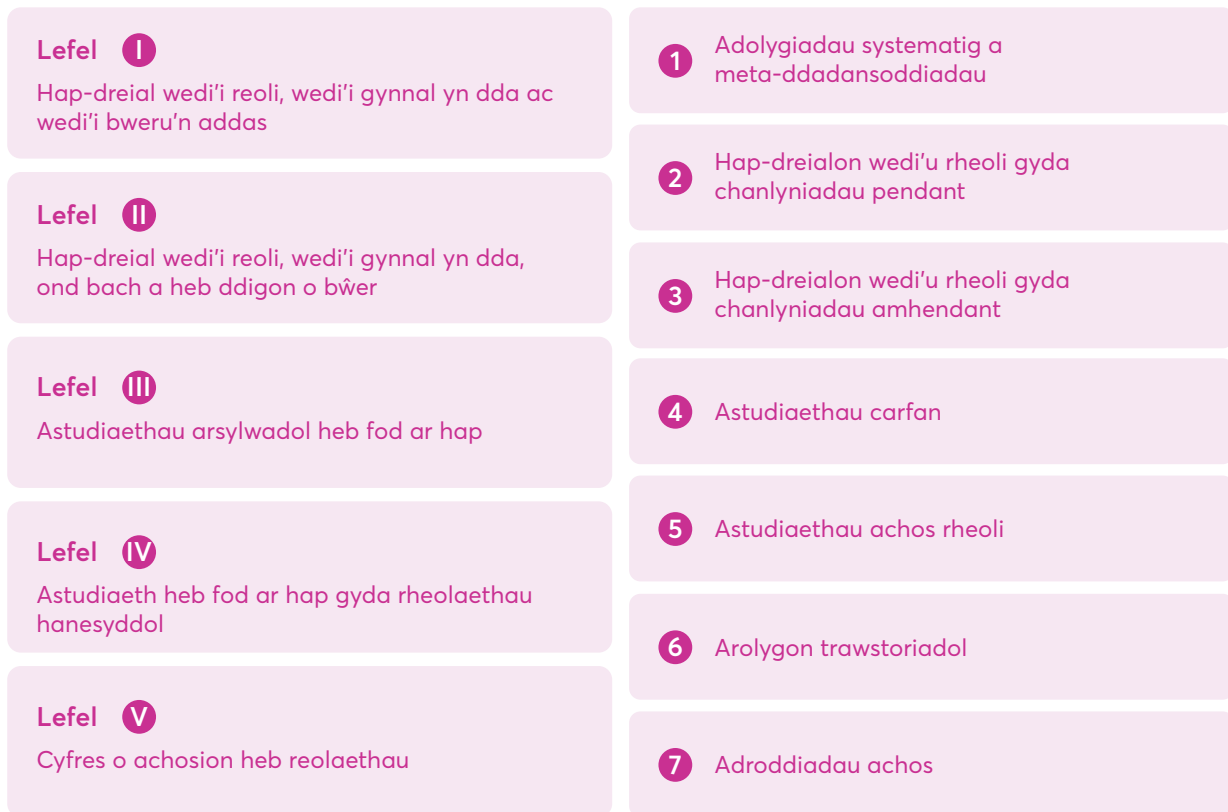
(Evaluating service transformation, blog NatCen, 19 Tachwedd 2014).

## Defnyddio haenau uchaf, hierarchaeth a safonau tystiolaeth

Ffordd arall o ystyried pwnc ansawdd yw defnyddio safonau ffurfiol o dystiolaeth – yn arbennig wrth edrych ar gwestiynau am 'beth sy'n gweithio' mewn polisi ac arfer cymdeithasol.

Gallai'r rhain fod ar ffurf hierarchaeth, lle mae dyluniadau ymchwil cryfach megis hap-dreialon wedi'u rheoli ac adolygiadau systematig ar y brig ar gyfer cwestiynau 'achosol' – a dulliau eraill, megis astudiaethau achos, ar y gwaelod.<sup>86</sup>

**Ffigur C2: Dwy ddelwedd o hierarchaeth tystiolaeth wedi'u symleiddio, yn seiliedig ar ddyfyniad yr astudiaeth**



Ffynhonnell: Bagshaw a Bellomo 2008, t.2.

Ffynhonnell: Petticrew a Roberts 2003, t.527

Caiff fersiynau gwahanol o hierarchaeth tystiolaeth eu defnyddio gan nifer o grwpiau gwerthuso ymchwil a chyrrff cymeradwyo o amgylch y byd, ac maent yn hynod gyffredin ym maes gofal iechyd. Mae'r safonau tystiolaeth hyn hefyd wedi bod yn tyfu ym maes polisi cymdeithasol y DU. Mae'r cyrrff a fu'n eu defnyddio yn cynnwys Public Health England,<sup>87</sup> Centre for Social Action – Nesta/Swyddfa'r Cabinet,<sup>88</sup> a Project Oracle yn Llundain,<sup>89</sup> sef prosiect i leihau troseddau ieuenctid.

### Ffigur C3: Safonau Tystiolaeth Nesta



Datblygodd Nesta ei fframwaith Safonau Tystiolaeth yn 2012 i asesu ansawdd y dystiolaeth ar effaith y mentrau cymdeithasol y buddsoddom ynddynt. Mae buddsoddi mewn effaith yn fusnes mawr ac mae wedi cael cefnogaeth sylweddol gan lywodraeth y DU. Ond ar gyfer cronfa buddsoddi Nesta, roedd y pryder yn ymwneud â sut y gallem fod yn sicr bod yr arloesiadau newydd wir yn cael effaith gadarnhaol. A oedd hi'n bosibl eu bod yn gwneud niwed hyd yn oed? Roedd angen tystiolaeth gref i ddangos p'un a oedd cynnydd gwirioneddol yn cael ei wneud gyda'r buddsoddiadau.

Mae pum lefel o dystiolaeth yn fframwaith Nesta.<sup>90</sup> Mae'r rhain yn dechrau o'r lefel sylfaenol (lefel 1) lle gallwch ddisgrifio'r hyn a wnewch a pham ei fod yn bwysig, mewn modd rhesymegol, clir a pherswadiol. Rydych wedyn yn symud fyny'r raddfa tuag at gasglu data arferol (lefel 2), ac wedyn ymlaen i lefelau uwch lle rydych chi'n defnyddio grwpiau cymharu (lefel 3), cyn creu tystiolaeth o ailadrodd a chynyddu maint (lefelau 4 a 5). Dylai lefel y dystiolaeth sydd ei hangen arnoch fod yn briodol ar gyfer ble'r ydych chi o ran datblygu polisi neu ymyriad. Ar gyfer Nesta, bu'n hynod ddefnyddiol i'r mentrau a ariannwn fynegi sut mae eu cynnyrch neu wasanaeth yn arwain at newid cymdeithasol cadarnhaol.<sup>91</sup>

Mae beirniaid yr hierarchaethau hyn yn dweud y gallant fod yn gyfyngiad sy'n rhy anhyblyg. Mae angen i benderfyniadau am ansawdd newid yn ôl y cyd-destun ac ni ddylai fod un maint yn addas i bawb. Yn ddiddorol, mae'r Sefydliad Cenedlaethol dros Ragoriaeth mewn Gofal ac Iechyd (NICE) – un o ganolfannau What Works llywodraeth y DU (gweler tudalen 39) – bellach wedi gollwng ei hierarchaeth o dystiolaeth. Bu eu cyn-gadeirydd, Michael Rawlins, yn agored feirniadol o ddilyn hierarchaethau tystiolaeth yn slafaid.<sup>92</sup> Roedd yn ffafrio ffordd fwy craff o benderfynu beth sy'n briodol, yn hytrach na bod hap-dreialon wedi'u rheoli ac adolygiadau systematig bob amser yn drech na'r gweddill.<sup>93</sup>

Waeth beth fo rhinweddau'r beirniadaethau hyn,<sup>94</sup> o leiaf mae'r hierarchaethau, y fframweithiau a'r egwyddorion hyn yn cynnig strwythur i wirio honiadau'r dystiolaeth. Mae model Nesta wedi ei roi ar waith a'i brofi. Maen nhw'n cynnig strwythur cymharol hawdd i'w deall ac annhechnegol ar gyfer 'y daith dystiolaeth' y mae angen i nifer o sefydliadau ei dilyn; o elfennau sylfaenol 'damcaniaeth newid' i astudiaethau ailadrodd lluosog. Mae safonau tystiolaeth yn cynnig ffordd i asesu cryfder honiadau o effeithlonrwydd, ac o osgoi meddwl bod pob darn o dystiolaeth yn gyfartal.

Mae'r safonau hyn hefyd yn gymesur ac yn ystyried ymyriadau cam cynnar. Nid oes yn rhaid i brosiect newydd ddangos ei effaith ar unwaith – na chomisiynu gwerthusiad drud o'r diwrnod cyntaf. Mae ar syniadau newydd angen amser i dyfu. Ond mae cyrraedd lefelau uwch o dystiolaeth (lefelau 4 a 5) yn arwain at ofyn cwestiwn am y ffordd rydych chi'n barnu cryfder y corff o dystiolaeth, nid astudiaeth unigol yn unig – rhywbeth y byddwn yn edrych arno nesaf gydag adolygiadau systematig.

## Osgoi dethol drwy ddefnyddio adolygiadau systematig

Mae'n bwysig i'r rhai sy'n gwneud penderfyniadau feddwl am ansawdd 'sylfaen' y dystiolaeth – nid darnau unigol o dystiolaeth, ond casgliadau o ymchwil wedi'u cyfuno.

Mae llawer o'r farn ar ansawdd sydd wedi'i nodi uchod wedi canolbwyntio ar astudiaethau unigol mewn 'ymchwil gynradd' – hynny yw, ymchwil wreiddiol megis arbrofion, arolygon ac ati. Mae'n rhaid i ni hefyd feddwl am 'ymchwil eilaidd', ac am grynhoi a chasglu ymchwil bresennol ynghyd – tasg sy'n debygol o fod yn dasg ddesg. Rydyn ni'n llawer mwy tebygol o ddod o hyd i arweinydd elusen neu swyddog llywodraeth yn ymgymryd â'r math hwn o waith eilaidd.

Perygl cyffredin yw cynnal 'adolygiad llenyddiaeth'. Mae'r rhain yn hynod amheus. Fel y dywedodd Ben Goldacre, maen nhw'n annog temtasiwn i ddethol – yn ymwybodol neu'n anymwybodol – eich hoff rannau o dystiolaeth er mwyn cyd-fynd â'ch dewis gasgliadau:

*"Yn hytrach na stelcian drwy'r llenyddiaeth ymchwil, gan fynd ati'n ymwybodol neu'n anymwybodol i ddethol papurau yma ac acw sy'n cefnogi [ein] credoau sy'n bodoli eisoes, [rydym] yn defnyddio agwedd wyddonol systematig ar gyfer y broses benodol o edrych ar dystiolaeth wyddonol, gan sicrhau bod [ein] tystiolaeth mor gyflawn â phosibl ac mor gynrychioliadol â phosibl o'r holl dystiolaeth a wnaed erioed."*

Ben Goldacre (2012) *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients*, (London: Fourth Estate).

Rydym wedi gweld yn yr adran flaenorol ar dueddiadau gwybyddol y gallwn syrthio'n anymwybodol i'r fagl o edrych am dystiolaeth sy'n cyd-fynd â'n credoau. Felly mae angen inni fod yn ofalus iawn ein bod yn casglu'r holl ymchwil sydd ar gael, hyd yn oed yr ymchwil sydd â ffeithiau anghyfleus.

Un ffordd o osgoi'r arfer hon o ddethol yw defnyddio'r hyn a alwn yn adolygiadau systematig.<sup>95</sup> Nod y dulliau hyn yw bod yn drwyadl, a chanfod gymaint o ymchwil sy'n berthnasol i'ch cwestiwn ag sy'n bosibl. Maen nhw'n defnyddio dulliau clir i nodi beth y gellir ei ddweud yn ddibynadwy ar sail yr astudiaethau hyn. Gall yr adolygiadau hyn sgrinio astudiaethau am ansawdd – ar sail dyluniad a dulliau'r ymchwil.



Mae dulliau systematig hefyd o werth am eu bod yn glir ynglŷn â'r ffordd y gwnaethant chwilio am ymchwil – felly, o ran egwyddor o leiaf,<sup>96</sup> gallai eraill efelychu'r adolygiad systematig, sy'n wahanol i'r 'stelcian' mae Ben Goldacre yn ei ddisgrifio uchod.

## Nodweddion adolygiad systematig

- Mae'n cynnwys set o amcanion wedi'u nodi'n glir gyda meini prawf cymhwysedd wedi'u diffinio ymlaen llaw ar gyfer astudiaethau;
- Mae'n cynnwys methodoleg glir, y gellir ei hatgynhyrchu;
- Mae'n defnyddio dull chwilio systematig sy'n ceisio nodi'r holl astudiaethau sy'n cyrraedd y meini prawf cymhwysedd;
- Mae'n cynnwys asesiad ffurfiol ar ddilysrwydd canfyddiadau'r astudiaethau sydd wedi'u cynnwys;
- Mae'n cynhyrchu cyflwyniad, a chyfosodiad, systematig o nodweddion a chanfyddiadau'r astudiaethau sydd wedi'u cynnwys.

Ffynhonnell: Trysorlys Ei Mawrhydi (2011) *The Magenta Book; Guidance for Evaluation*.

## Meta-ddadansoddiad

Yn aml, mae meta-ddadansoddiad yn rhan bwysig o adolygiadau systematig ac mae fel arfer yn cynnwys dwyn swm mawr o ganfyddiadau ymchwil ynghyd – ond gan ddefnyddio dadansoddiad ystadegol. Yn aml, mae'n golygu creu cronfa o feintiau cyfartalog yr effaith a amcangyfrifir ym mhob astudiaeth ymchwil unigol.

Mae'n debyg fod meta-ddadansoddiad yn fwyaf cyfarwydd am gyfuno canlyniadau hap-dreialon wedi'u rheoli, ond mae hefyd yn cael ei ddefnyddio'n gyffredin ar ddata nad yw wedi'i ddewis ar hap – o astudiaethau cynradd sy'n defnyddio rheoli achos, i ddyluniadau trawstoriadol a charfan (gweler Tabl C1 ar dudalen 22).

## Gwahaniaethu rhwng meta-ddadansoddi, meta-werthuso ac adolygiadau systematig

Mae gwahaniaeth pwysig rhwng adolygiadau systematig a meta-ddadansoddi, ac ni ddylid cymysgu rhyngddynt. Mae adolygiad systematig yn defnyddio dulliau systematig ar gyfer adolygu llenyddiaeth. Gellid ei hystyried yn broses 'feta lefel' (ynghylch neu du hwnt) gan mai ymchwil ar ymchwil ydyw. Felly gellid defnyddio'r term 'meta-ddadansoddi' i ddisgrifio'r holl adolygiadau systematig, ond yn ymarferol caiff y term ei ddefnyddio ar gyfer meta-ddadansoddi ystadegol yn unig, lle caiff ystadegau eu defnyddio i gyfosod y data rhifiadol o astudiaethau gwahanol.

Hefyd, mae'r term 'meta-werthusio' (Scriven 1969) a all olygu 'gwerthuso gwerthusiadau', a gall gynnwys cyfosod astudiaethau gwerthuso, ond mae ganddo ystyron ehangach (Scriven 1969, Gough et al. 2012b).<sup>97</sup> I gael disgrifiad hirach o 'feta-werthuso' a sut mae'n wahanol i fathau eraill o adolygiadau, gweler y *Magenta Book*.

## Adolygiadau brys

Ond gall adolygiadau systematig a meta-ddadansoddi gymryd amser, ac nid yw hyn yn dda pan fo gofyn gwneud penderfyniadau ar frys. Mae'n debyg na allwch aros am chwe mis i gwblhau adolygiad mewn pryd ar gyfer araith fawr yfory, dyddiad cau gwneud cais am gyllid, neu femo i weinidog ar bolisi newydd. Fodd bynnag, mae nifer o'r adolygiadau systematig hyn 'ar y silff' eisoes, felly nid oes angen i chi ddechrau o'r dechrau. Maen nhw ar gael i bawb mewn llyfrgelloedd ar-lein megis y rhai a guradwyd gan Cochrane and Campbell Collaborations, sy'n cael eu trafod yn Adran D ar ble i ddod o hyd i dystiolaeth.

Er hynny, efallai y teimlwch yn rhwystredig wrth fethu â dod o hyd i'r hyn y chwiliwch amdano ar y gwefannau hynny. Felly eich opsiwn gorau yw comisiynu fersiwn lai o'r adolygiad systematig,<sup>98</sup> fel asesiadau tystiolaeth cyflym. Nod asesiad tystiolaeth cyflym yw adolygu tystiolaeth ar frys, heb gyfaddawdu'n ormodol ar ansawdd (gweler Adran D i wybod mwy am asesiadau tystiolaeth cyflym).<sup>99</sup>

Fodd bynnag, gan fod y cwestiynau rydyn ni am eu hateb yn awr yn debygol iawn o godi eto mewn blynyddoedd i ddod, ar y lefel sefydliadol mae'n gwneud synnwyr comisiynu adolygiadau o ansawdd uchel a gaiff eu diweddarau'n rheolaidd ac sydd ar gael i bawb. Dyma'n union yw rôl y llyfrgelloedd ar-lein sy'n cael sylw yn Adran D.

Noder, fodd bynnag, nad yw ffyrdd mwy ffurfiol o ddwyn ymchwil ynghyd bob amser yn gyfystyr ag ansawdd. Mae barnu ansawdd eich adolygiad tystiolaeth yn dal i fod yn anodd. Mewn gwirionedd, nid yw ansawdd eich cyfodiad damaid yn well nag ansawdd yr astudiaethau y mae'n seiliedig arnynt.

## Pwysigrwydd ailadrodd a chadarnhau

Mae pethau eraill sydd angen eu hystyried wrth asesu ansawdd crynodeb ymchwil. Er enghraifft, faint o astudiaethau sydd angen eu cynnwys er mwyn i chi fod yn fodlon bod corff cryf o dystiolaeth yn bodoli? Efallai mai ychydig o astudiaethau yn unig os ydyn nhw'n dda iawn? Neu ddwsinau efallai, neu gannoedd? Mae Gwiriad Tystiolaeth diweddar gan yr Adran Addysg<sup>100</sup> yn crybwyll 12 o adroddiadau Ofsted ar ysgolion cynradd sy'n awgrymu bod perfformiad cryf yn gysylltiedig ag addysgu ffoneg. Ond mae tua 16,000 o ysgolion cynradd a ariennir gan y wladwriaeth yn Lloegr, felly nid yw 12 yn ymddangos yn llawer. Ac mae hynny hefyd yn osgoi'r cwestiwn difrifol arall o ran p'un a allwn ddweud mewn gwirionedd a yw arolygiadau Ofsted yn cyfrif fel tystiolaeth gref o effaith.

Y gwirionedd yw nad oes rhif hud o ran nifer o astudiaethau. Ac eto, ni allwn anwybyddu'r ffaith fod maint corff y dystiolaeth yn bwysig: mae cryfder mewn niferoedd, ac mae'n rhaid wrth ailadrodd a chadarnhau. Mae hyd yn oed angen i astudiaethau sydd wedi'u canmol droeon gael eu hailadrodd drosodd a thro. Fe wnaeth Amgen, cwmni cyffuriau o Galiffornia, geisio efelychu 53 o astudiaethau canser nodedig. Chwech yn unig (11 y cant) o'r astudiaethau pwysig hyn a efelychwyd yn y gwaith, a gyhoeddwyd yn Nature.<sup>101</sup> Mae hyn yn destun pryder mawr gan fod yr astudiaethau wedi dylanwadu ar gwmnïau cyffuriau a thriniaethau canser yn fyd-eang.

## 'Adolygu adolygiadau' – sut i farnu cyrff o dystiolaeth o ansawdd da

Yn sicr, dylem fod yn wyladwrus o astudiaethau unigol – am gyffuriau rhyfeddol newydd ar gyfer Alzheimer's sy'n llenwi tudalen flaen y *Daily Mail*, er enghraifft – nad ydynt yn cael eu hefelychu byth.<sup>102</sup> Ond nid maint y dystiolaeth yw'r unig beth i'w ystyried. Mae How to Note ar farnu cryfder tystiolaeth a gynhyrchwyd gan DFID yn rhestru pedwar peth i'w ystyried wrth wirio cyrff o dystiolaeth:

1. Ansawdd (technegol) yr astudiaethau sy'n rhan o'r corff o dystiolaeth (neu i ba raddau'r aethpwyd i'r afael â'r perygl o duedd);
2. Maint y corff o dystiolaeth;
3. Cyd-destun y dystiolaeth;
4. Cysondeb y canfyddiadau a gynhyrchwyd gan astudiaethau sy'n rhan o'r corff o dystiolaeth.

Ffynhonnell: DFID (2014) *How to Note: Assessing the Strength of Evidence*

Mae barnu'r holl feini prawf hyn bob amser am fod yn oddrychol – ac yn berthnasol i gyd-destun cwestiwn y polisi. Mae'n anodd iawn rhoi cyngor cyffredinol ar y corff iawn o dystiolaeth ar gyfer unrhyw bolisi. Ymagwedd ddiddorol arall yw dwyn casgliadau o adolygiadau systematig ynghyd yn unig, a pheidio â rhidyllu drwy astudiaethau unigol yn unig – hynny yw, 'adolygiad o adolygiadau'. Wrth gwrs, efallai mai pan fydd adolygiadau systematig yn bodoli'n barod yn unig y bydd hyn yn gweithio – ac mewn rhai sectorau, ychydig o'r rhain sydd ar gael. Mae model da i'w gael gan y Sefydliad Gwaddol Addysgol – elusen annibynnol a 'What Works Centre' a ariennir gan yr Adran Addysg ac Ymddiriedolaeth Sutton, a drafodir yn Adran D.

Mae rhestr o ganllawiau manwl eraill ar ansawdd ymchwil i'w gweld yn Atodiad 3.

### Negeseuon allweddol o Adran C Rhan 2

- I ddod o hyd i dystiolaeth y gallwch ymddiried ynnddi **edrychwch am ymchwil sydd wedi'i hadolygu gan gymheiriaid**. Ond noder nad yw adolygu gan gymheiriaid yn berffaith o bell ffordd. Gallai fod yn destun tueddiadau anymwybodol fel 'corlannu', neu duedd cyhoeddi tuag at ganlyniadau cadarnhaol, neu dwyll academaidd hyd yn oed.
- Wrth edrych ar gwestiynau ynghylch effaith a 'beth sy'n gweithio', defnyddiwch y **fframweithiau a safonau tystiolaeth ffurfiol** megis y rhai a ddefnyddir gan Nesta, Project Oracle ac eraill.
- Nid yw un astudiaeth byth yn ddigon ym maes polisi cymdeithasol. **Dylech osgoi gwneud penderfyniadau'n seiliedig ar astudiaethau unigol, ac edrych am efelychiadau amrywiol**. Mae cryfder mewn niferoedd, ac mae'n rhaid wrth ailadrodd a chadarnhau. Mae hyd yn oed angen i astudiaethau sydd wedi'u canmol droeon gael eu hailadrodd drosodd a thro.
- Peidiwch â gwneud 'adolygiad llenyddiaeth' o ymchwil, gan ei fod yn annog dethol tystiolaeth – yn ymwybodol neu'n anymwybodol. **Defnyddiwch adolygiadau systematig**, sydd â'r nod o fod yn drylwyr ac a all sgrinio ansawdd – fel arfer ar sail dyluniad a dulliau'r ymchwil.
- Os nad oes gennych amser i wneud adolygiad systematig, neu'n methu dod o hyd i un wedi'i gwblhau sy'n ateb eich cwestiwn, ewch ati i gomisiynu **asesiad tystiolaeth cyflym**.

# Adran D

## Ble dylech chi edrych am dystiolaeth?

Mae'r adran hon yn cynnig cyngor ac adnoddau i'ch helpu i ddod o hyd i'r dystiolaeth gywir i gefnogi eich achos.

The screenshot shows a Guardian article from October 24, 2013. The headline is "Greg Hunt uses Wikipedia research to dismiss climate change-bushfires link". The sub-headline reads: "Environment minister warns against 'politicising the science', saying bushfires are simply part of 'the Australian experience'". The author is Helen Davidson (@heldavidson). The article includes a photo of Greg Hunt speaking at a podium. A hashtag #GregHuntResearch is visible on the right side of the article. The article's share statistics show 6,500 shares and 322 comments.

Roedd yn hynod o onest i gyn Weinidog yr Amgylchedd yn Awstralia, Greg Hunt, gyfaddef ei fod wedi cael ei dystiolaeth o bori ar y we. "Edrychais ar beth oedd gan Wikipedia i'w ddweud," oedd ei ddatl ar BBC World Service pan ofynwyd iddo am safbwyntiau ei lywodraeth ar danau llwyni. Aeth sylwadau Mr Hunt dros y byd. Cafodd ei wawdio ar Twitter, gyda'i ddatganiad yn arwain at greu'r hashnod #GregHuntResearch. I atgyfnerthu'r pwynt, gwnaeth pobl ddrygonus Wikipedia ddiweddarau ei dudalen Wikipedia bersonol gan gynnwys nodyn: "was quoted as saying he uses Wikipedia for important policy research".

Byddech yn gobeithio y byddai ceidwaid polisi eraill yn edrych am ymchwil fwy cadarn na Wikipedia, ond nid yw'r gwleidydd o Awstralia ar ei ben ei hun. Dangosodd arolwg diweddar<sup>103</sup> gan Ymddiriedolaeth Carnegie UK fod y rhan fwyaf o lunwyr polisi yn cael eu tystiolaeth drwy'r rhyngwyd, papurau newydd a newyddion wedi'i ddarlledu.

Collwn gyfle fan hyn. Yn y man lleiaf, gallai Greg Hunt fod wedi chwilio ar beiriant chwilio hollbresennol **Google Scholar**. Mae amrywiaeth eang o adnoddau ymchwil dibynadwy ar-lein, ac mae nifer ohonynt am ddim ac yn hawdd i'w defnyddio. Bydd y bennod hon yn cyfeirio at ychydig yn unig y credwn y byddant yn ddefnyddiol i unrhyw lunwyr polisi, arweinwyr elusennau a gweithwyr proffesiynol rheng flaen.

Ffigur D1: Ecosystem Tystiolaeth y DU ar gyfer polisi cymdeithasol



### Canolfannau What Works y DU mewn polisi cymdeithasol (a pholisi economaidd lleol)

Mae lansio chwech o ganolfannau 'What Works' yn 2013 wedi trawsnewid tirwedd tystiolaeth ar gyfer polisi ac arfer cymdeithasol yn y DU.<sup>104</sup> Eu nod yw gwella'r ffordd mae llywodraeth a sefydliadau eraill yn creu, rhannu a defnyddio tystiolaeth o ansawdd ac maen nhw'n cael eu goruchwyllo gan y What Works Network sy'n cael ei redeg o Swyddfa'r Cabinet, yn ogystal â llywodraethau'r Alban a Chymru. Nawr, mae naw canolfan – i'w gweld yn y 'selsigen' oren ganolig yn y ffigur uchod – yn rhoi sylw i feysydd fel trosedd, addysg a lles. Maen nhw'n cael eu hariannu gan y Cyngor Ymchwil Economaidd a Chymdeithasol, y Gronfa Loteri Fawr ac ystod eang o adrannau'r llywodraeth – gan gynnwys llywodraethau'r Alban a Chymru.

Ni allwn ddefnyddio'r esgus mwyach fod ymchwil yn rhy anodd i'w ganfod oherwydd ei fod yn cuddio y tu ôl i waliau talu academiaidd, ond mewn rhyw jargon annealladwy. P'un a ydych chi mewn elusen, llywodraeth ganolog neu awdurdod lleol, maen nhw'n cynnig cyfosiadau gweithredol a hygyrch o ymchwil a thystiolaeth. Blaenoriaeth benodol

yw sicrhau bod eu crynodebau ymchwil yn cael eu cyfleu mewn ffyrdd y gall y sawl sy'n gwneud penderfyniadau uniaethu â nhw – nid llunwyr polisi'n unig, ond comisiynwyr gwasanaethau neu ymarferwyr rheng flaen megis athrawon a swyddogion yr heddlu. Mae'n ddyddiau cynnar o hyd i rai o'r canolfannau What Works hyn, felly nid yw eu holl adnoddau wedi'u cwblhau eto. Ond, ar adeg cyhoeddi, mae'r canlynol yn adnoddau defnyddiol ac agored sy'n crynhoi'r ymchwil:

- **Pecyn Cymorth Dysgu ac Addysgu Ymddiriedolaeth Sutton / Sefydliad Gwaddol Addysgol a Phecyn Cymorth Blynyddoedd Cynnar y Sefydliad Gwaddol Addysgol** – tystiolaeth o'r effaith ar gyrhaeddiad disgyblion ar gyfer ymyriadau yn y dosbarth neu o amgylch yr ysgol, megis mentora neu gymryd rhan mewn chwaraeon.

Cynulleidfa: athrawon, rhieni, penaethiaid, gweithwyr proffesiynol y blynyddoedd cynnar, llywodraethwyr, yn ogystal ag unrhyw elusen, awdurdod lleol neu lunwyr polisi.

- **Canolfan What Works ar gyfer Twf Economaidd Lleol** – adolygiadau systematig ar faterion polisi twf economaidd lleol megis mynediad at gyllid, prentisiaethau, band llydan, hyfforddiant cyflogaeth, parthau menter a chluant.

Cynulleidfa: unrhyw un sy'n ymwneud â gwneud penderfyniadau polisi sy'n targedu twf economaidd – yn enwedig awdurdodau lleol, Partneriaethau Cyflogaeth Lleol, llywodraeth, a busnesau.

- **Pecyn Cymorth y Ganolfan What Works ar gyfer Lleihau Trosedd yn y Coleg Plismona** – mae dros 300 o adolygiadau systematig, sy'n mynd i'r afael â 60 o ymyriadau lleihau trosedd gwahanol, wedi'u nodi ac yn cael eu hychwanegu at y teclyn ar-lein dros amser, megis CCTV, goleuadau stryd ac ymweliadau â charchar i ddigalonni pobl ifanc rhag troseddu.

Cynulleidfa: yr heddlu a gweithwyr proffesiynol lleihau trosedd, llunwyr polisi, comisiynwyr gwasanaethau cyhoeddus, elusennau a grwpiau gwirfoddol sydd â diddordeb mewn rhaglenni ac ymyriadau lleihau trosedd.

The screenshot shows the 'What Works Crime Reduction Toolkit' website. The main heading is 'Crime Reduction Toolkit'. Below this, there is a table of interventions. The table has six columns: 'Intervention', 'Impact on crime', 'How it works', 'Where it works', 'How to do it', and 'What it costs'. Each row represents an intervention with corresponding icons and quality indicators. A sidebar on the left includes filters for 'Our effect scale', 'Our quality scale', and 'Impact on Crime'.

Intervention	Impact on crime	How it works	Where it works	How to do it	What it costs
	Effect	Mechanism	Moderator	Implementation	Economic cost
Alcohol ignition interlock	✓✓	⚙️	📍	❓	£
Alcohol tax and price policies	✓✓	⚙️	📍	❓	£
<b>NEW</b> Alternative education programmes	X✓	⚙️	📍	❓	£
CCTV	✓✓	⚙️	📍	❓	£
Cognitive Behavioural Therapy (CBT)	✓✓	⚙️	📍	❓	£



- **Canllaw a Llyfrgell Rhaglenni Ymyriad Cynnar, y Sefydliad Ymyriad Cynnar** – mae'n mynd i'r afael â thystiolaeth ar dros 50 o raglenni ymyriad cynnar. Mae'n ceisio annog datblygiad cadarnhaol ac yn ymdrin ag amrywiaeth o broblemau, gan gynnwys cam-drin ac esgeuluso, cam-drin sylweddau a phroblemau iechyd meddwl. Lluniwyd y canllaw i roi cyngor i weithwyr proffesiynol sy'n datblygu systemau a rhaglenni ymyriad cynnar.

**Cynulleidfa:** comisiynwyr gwasanaethau, llunwyr polisi ac ymarferwyr ar draws maes polisi cymdeithasol..

- **Adolygiadau tystiolaeth Sefydliad Polisi Cyhoeddus Cymru** – mae Sefydliad Polisi Cyhoeddus Cymru yn darparu dadansoddiad a chngor cyflym i weinidogion llywodraeth Cymru. Y mae hefyd yn ymgymryd â gwaith ar rai o'r prif heriau strategol sy'n wynebu llywodraeth Cymru, megis mynd i'r afael â thlodi. Mae ei wefan yn cynnwys dolenni i ddwsinau o grynodedau ymchwil sydd wedi'u cwblhau ac sy'n hygyrch i bawb gan gynnwys adolygiadau tystiolaeth cyflym, cyngor a dadansoddiad arbenigol gan y Sefydliad, yn amrywio o gynllunio trafniadaeth strategol i effaith dyled yng Nghymru.

**Cynulleidfa:** gweinidogion llywodraeth Cymru, eraill sydd â diddordeb mewn llunio polisi cymdeithasol yng Nghymru.

Cadwch lygaid am grynodedau ymchwil ar-lein yn y dyfodol gan ganolfannau eraill What Works:

- What Works in Tackling Poverty (dan arweiniad Sefydliad Polisi Cyhoeddus Cymru – sydd hefyd yn cynnal Canolfan What Works Cymru).
- What Works Scotland Evidence Bank.
- What Works Centre for Wellbeing.
- Centre for Ageing Better.

## Gwaith NICE mewn iechyd a gofal cymdeithasol

Arweinydd yng nghanolfannau What Works y DU yw'r Sefydliad Cenedlaethol dros Ragoriaeth mewn Iechyd a Gofal (NICE). Mae'r holl arweiniad a gynigir gan NICE yn seiliedig ar y dystiolaeth ymchwil orau sydd ar gael – yn ogystal â chyfraniad gan gleifion drwy gyrff tebyg i 'reithgor dinasyddion', a chlinigwyr.<sup>105</sup> Mae NICE hefyd yn cynnig dau adnodd defnyddiol arall.

- **NHS Evidence:** cronfa ddata ddefnyddiol sy'n rhoi canllawiau ac ymchwil dealladwy parod i'r feddygfa neu'r ward ysbyty. Mae'n adnodd rhyfeddol sy'n amrywio o gronfeydd data ar feddyginiaethau sy'n agored ac am ddim i ddolenni (ar gyfer gweithwyr y GIG yn unig) i'r prif gyfnodolion a chronfeydd data llyfryddol.
- **UK DUETS:** talfyriad ydyw o **Database of Uncertainties about the Effects of Treatments**, ac mae'n tynnu sylw at gwestiwn difrifol sy'n dal i fodoli ynghylch p'un a yw triniaethau'n gweithio ai peidio. Mae'n seiliedig ar beth mae cleifion a chlinigwyr yn ei ddweud, yn ogystal â'r ymchwil gynradd a'r adolygiadau systematig gorau sydd ar gael. Mae'r gronfa ddata'n agored am fylchau yn ein dealltwriaeth ac yn cydnabod bod anwybyddu ansicrwydd o ran triniaethau yn rhoi cleifion mewn perygl yn y DU.



## Cronfeydd ymchwil eraill ar gyfer polisi ac arfer cymdeithasol

Nid yw'r rhan fwyaf o gronfeydd mor addas i ymarferwyr ag NHS Evidence, ac maent wedi'u targedu mwy at ymchwilwyr arbenigol. Ond gall rhai o'r cronfeydd data eraill hyn fod yn rhyfeddol o hawdd i'w defnyddio:

- Mae gan yr Evidence for Policy and Practice Information Co-ordinating Centre yng Ngholeg Prifysgol Llundain amrywiaeth o gronfeydd data sydd ar gael yn rhad ac am ddim ar-lein, megis y Database of Education Research, gyda 1,200 o gofnodion o adolygiadau systematig. Gallwch chwilio ei holl adolygiadau systematig yn ôl gair allweddol.
- Hefyd, mae cronfeydd data gwyddonol mawr o gyfnodolion, fel PubMed ar gyfer astudiaethau mewn iechedd cyhoeddus neu JSTOR ar gyfer astudiaethau mewn economeg, cymdeithaseg a pholisi cyhoeddus – er bod hwn wedi'i dargedu at academyddion yn bennaf. Sefydliad nid-er-elw yw JSTOR, ond y mae'n codi ffi i dalu ei gostau a chaiff ei ddefnyddio'n bennaf mewn prifysgolion, llyfrgelloedd cyhoeddus ac ysgolion. Os ydych chi'n gweithio yn llywodraeth ganolog y DU neu awdurdod lleol, efallai y bydd eich adran yn tanysgrifio a gall defnyddwyr sy'n 'cerdded i mewn' fanteisio ar dros 2,000 o gyfnodolion academaidd.
- Os oes gennych chi gysylltiadau agos gyda'ch hen brifysgol yn y DU, cofiwch fod gan gyn-fyfyrwyr yn aml fynediad o bell at gronfeydd data llyfrgelloedd eu cyn brifysgol, megis adnodd Cyn-fyfyrwyr Prifysgol Rhydychen.

## Adolygiadau systematig eraill y tu hwnt i ganolfannau What Works

Yn yr adran flaenorol, argymhellwyd defnyddio adolygiadau systematig, yn hytrach nag astudiaethau unigol yn unig. Er mai drud yw eu comisiynu o'r cychwyn cyntaf, mae llawer o adolygiadau wedi'u cwblhau sy'n gymharol hygyrch.

### Polisi cymdeithasol cyffredinol

Mae gan ganolfan **Evidence for Policy and Practice Information** yng Ngholeg Prifysgol Llundain y cyfeiriwyd ati uchod 'lyfrgell dystiolaeth' o adolygiadau systematig y gall unrhyw un fanteisio arnynt, er nad yw'r ystod yn gynhwysfawr gan ei bod yn cynnwys yr astudiaethau y bu'r Ganolfan yn rhan ohonynt yn unig.

Gwefan fwy cynhwysfawr ar gyfer rhestrau o adolygiadau systematig ym maes polisi cymdeithasol yw'r **Campbell Collaboration Library**. Mae'n cwmpasu addysg, trosedd a chyfiawnder, lles cymdeithasol a datblygiad rhyngwladol. A hithau wedi'i rheoli gan Ysgrifenyddiaeth, mae'r Llyfrgell wedi'i rhannu'n Grwpiau Cydlynau sy'n rheoli proses adolygu cymheiriaid drylwyr ar gyfer adolygiadau systematig o ansawdd uchel. Gall yr adolygiadau fod ar destunau hynod bynciol sy'n boblogaidd ymysg rhai sy'n gwneud penderfyniadau, megis ei adolygiadau o 'ymwybyddiaeth ofalgar' wrth leihau straen sydd wedi'i lawrlwytho 10,000 o weithiau.

## Addysg

Adnodd rhad ac am ddim arall ar wahân i Ganolfan What Works y Sefydliad Gwaddol Addysgol yw gwefan **Evidence for Impact** sy'n rhoi system sgorio tystiolaeth syml ar gyfer rhaglenni sy'n ymwneud ag ysgolion, ynghyd â chrynodebau tystiolaeth byr. Mae'n cynnwys cronfa ddata gynhwysfawr o raglenni sydd ar gael yn y DU, gan gynnwys manylion am eu heffeithiolrwydd a'u cost, ynghyd â dolenni i ddarparwyr ac arbenigwyr a all gynnig cefnogaeth bellach. Gellir chwilio'n hawdd yn y gronfa yn ôl cyfnod allweddol, maes pwnc a grŵp wedi'i dargedu, fel bod y canlyniadau wedi'u teilwra i anghenion penodol eich dosbarth neu ysgol.

## Gwasanaethau plant

Mae gan Uned Ymchwil Cymdeithasol Dartington siop un-stop ar gyfer comisiynwyr gwasanaethau plant o'r enw **Investing in Children**. Mae'n dwyn tystiolaeth ynghyd o ran 'beth sy'n gweithio' gyda chostau a manteision economaidd 100 o ymyriadau gwahanol sy'n cystadlu â'i gilydd ym maes gwasanaethau plant.

## Datblygu rhyngwladol

O ran datblygu rhyngwladol, mae cronfa ddata **3ie** yn mynd i'r afael ag astudiaethau cynradd ac adolygiadau systematig o effeithiolrwydd ymyriadau cymdeithasol ac economaidd mewn gwledydd incwm isel ac incwm canolig.

Yn hytrach na mynd drwy bob un o'r pyrth neu wefannau uchod, efallai yr ystyriwch fynd yn syth i Google Scholar. Fel awgrym, i ddod o hyd i adolygiadau systematig defnyddiwrch y termau chwilio 'subject' YN OGYSTAL Â 'systematic review' e.e. 'mindfulness' A 'systematic review'. Os ydych chi eisiau adolygiadau gan sefydliadau penodol gallwch ehangu'r llinyn chwilio i nodi hynny, e.e. 'mindfulness' A 'systematic review' A 'Campbell' neu sefydliad arall sy'n cynhyrchu adolygiadau.

## Siop un-stop yw'r stop olaf

Mae nifer o siopau un-stop sy'n crynhoi adnoddau tystiolaeth sy'n ymwneud â thestun penodol, ond y perygl yw y byddant yn disgyn i fagl **Portal Proliferation Syndrome**. Hynny yw, ni all cynhyrchwyr ymchwil wrthod temtasiwn sefydlu canolfan, porth, llwyfan neu fynedfa arall i geisio pecynnu'r wybodaeth oll mewn un man. Gallai cronfeydd data o gronfeydd data wneud dim ond drysu defnyddwyr.

Er hyn, mae creu cronfa o wybodaeth ar ffurf parod i'w defnyddio ar gyfer proffesiynau polisi cymdeithasol yn sicr yn nod da. Ac o'i wneud yn dda, gall dwyn cronfeydd data ynghyd mewn un lle helpu i leihau dryswch. Mae rhai adnoddau digidol da iawn i'w cael ar gyfer cyfod. Ymatebodd siop un-stop Social Policy and Practice ar gyfer ymchwil ym maes iechyd a pholisi cymdeithasol i'r swm mawr o adnoddau gwybodaeth drwy gydblethu pedair cronfa ddata ymchwil genedlaethol flaenorol sy'n berthnasol i bolisi: ChildData, AgeInfo, Planex a Social Care Online.

Mae Social Policy and Practice yn ddefnyddiol ar gyfer unrhyw weithiwr proffesiynol sy'n gweithio ym maes gofal cymdeithasol neu waith cymdeithasol nad oes modd iddynt ddefnyddio llyfrgell prifysgol yn hawdd. Yn ogystal â helpu i gael mynediad at 600 o gyfnodolion o'r DU a chyfnodolion rhyngwladol, mae hefyd yn cynnwys 'llynyddiaeth lwyd' adroddiadau a chanllawiau llywodraeth neu felinau trafod a fydd yn helpu'r rhai hynny sy'n gweithio ym maes gofal cymdeithasol. Mae canllaw cryno i'r cronfeydd data ymchwil hyn yn y DU i'w gweld ar wefan y Gynhrai ar gyfer Tystiolaeth Ddefnyddiol.<sup>106</sup>

## Crynodebau bach, byr, a bachog

Gall rhai o adroddiadau Campbell Collaboration fod hyd at 100 tudalen o hyd, ond bydd angen rhywbeth byrrach ar rywun sy'n gwneud penderfyniadau ac sy'n brin o amser. Gallai crynodebau ddweud tipyn wrthyf, ac mae digonedd o **gyfnodolion academaidd** sy'n rhoi crynodebau – os nad erthyglau cyfan – am ddim. Mae adolygiadau Campbell yn cynnwys 'crynodeb iaith glir', sy'n cael eu gwneud yn fwy hygyrch ar hyn o bryd.

Mae naratif darllenadwy ac anacademaidd eraill sy'n crynhoi adolygiadau o waith ymchwil, wedi'u hysgrifennu mewn ffordd sy'n addas i ymarferwyr (h.y. gan beidio â defnyddio jargon academaidd oni bai ei fod yn hanfodol), fel **crynodebau tystiolaeth**<sup>107</sup> Llywodraeth yr Alban 2011 ar yr hyn sy'n gweithio i atal pobl rhag ail-droseddu, ac wedyn **crynodeb o dystiolaeth**<sup>108</sup> Llywodraeth y DU 2013 ar yr un pwnc. Neu'r erthyglau pynciol gan academyddion ar wefan The Conversation, gan gynnwys adran **Hard Evidence**.

Mae **POSTNotes** seneddol yn grynodebau ymchwil PDF sydd hefyd yn fyr a hawdd eu deall, a gall unrhyw un eu gweld, nid Aelodau Seneddol yn unig. Er eu bod yn ymwneud â gwyddoniaeth a thechnoleg yn bennaf, mae ganddynt nifer gynyddol ym maes gwyddor gymdeithasol a pholisi cymdeithasol. Dylech danysgrifio i dderbyn **Evidence Briefings** rheolaidd gan y Cyngor Ymchwil Economaidd a Chymdeithasol; ac os oes gennych chi ddiddordeb ym maes polisi plant, mae Evidence Nuggets o'r What Works Centre for Children (bellach ar gau) yn dal i fod ar gael.<sup>109</sup>

## Gwasanaethau ymgynghori a desg gymorth – talu'r pris am dystiolaeth

Fodd bynnag, nid yw rhai siopau un-stop ar agor i bawb ac efallai y bydd angen tanygrifiad neu daliad. Ond os yw'r cynnyrch yn dda, onid yw'n werth buddsoddi? Efallai y bydd rhai ohonom yn ddigon ffodus i gael mynediad at lyfrgelloedd prifysgolion (a llyfrgellwyr da – arwyr di-glod byd tystiolaeth) er mwyn mynd y tu hwnt i waliau talu cyfnodolion ymchwil uchel eu parch. Ond i'r rhai ohonom na allwn wneud hynny (gan gynnwys awduron y canllaw hwn, er ein bod yn gweithio i elusen fawr yn y DU), a yw'n werth talu i rywun gasglu eich tystiolaeth?

Un enghraifft o ddarparwr sy'n codi ffi yw'r Idox Information Service. Mae wed bodoli ers 30 mlynedd bellach ac mae'n honni bod ganddo 200,000 o ddefnyddwyr. Yn ogystal â chael cronfa ddata o grynodebau, mae ymgynghorwyr wrth law i'ch helpu. Mae hyd yn oed ganddo wasanaeth llyfrgell hen ffasiwn – gallwch rentu llyfrau drwy'r post. Mae Idox yn honni mai nhw yw'r ffynhonnell fwyaf cyflawn a hygyrch o dystiolaeth ac ymchwil ar gyfer y sector cyhoeddus yn y DU, ond mae'n rhaid i chi dalu amdano.

Mae sefydliadau ymgynghoriaeth ymchwil eraill a all eich helpu i ddod o hyd i dystiolaeth – mae rhestr o'r rhain sydd hefyd yn aelodau o'r Gynghrair i'w gweld ar y wefan.

Un dull a allai fod yn ddefnyddiol yw asesiad tystiolaeth cyflym. Ei nod yw bod yn eithaf cynhwysfawr, ond yr anfantais yw nad yw'n drwyadl, gyda risg cysylltiedig o duedd.<sup>110</sup>

Mae'r tabl hwn gan y Sefydliad Cenedlaethol er Ymchwil i Addysg yn dangos beth gallech gael drwy gomisiynu asesiad tystiolaeth cyflym, neu fath arall o adolygiad cyflym megis astudiaeth gwmpasu:

Tabl D2:

Cynnyrch	Pam fyddwn i'n ei ddefnyddio?	Beth fydd yn ei roi i mi?	Faint o amser y bydd yn ei gymryd	Awgrym o hyd yr adroddiad
Adolygiad cyflym	Cael trosolwg cyflym o faes polisi neu arfer	Trosolwg o'r hyn a wyddom am bwnc penodol. Gall fod yn seiliedig ar lenyddiaeth ymchwil, polisi neu dystiolaeth arfer. Caiff y llenyddiaeth ei chanfod drwy chwilio'n gyfyngedig	1–2 fis	Adroddiad 10 tudalen
Adolygiad cwmpasu	Canfod p'un a oes digon o dystiolaeth i gynnal asesiad tystiolaeth cyflym neu adolygiad systematig llawn	Map o faint o dystiolaeth sydd ar gael am bwnc	1–2 fis	Adroddiad 10 tudalen
Adolygiad tystiolaeth cyflym	Cael cyfosodiad trylwyr o dystiolaeth i gyfarwyddo polisi neu arfer. Ateb cwestiynau am yr hyn sy'n effeithiol	Cyfosodiad cadarn o'r dystiolaeth, gyda negeseuon allweddol ac argymhellion. Caiff y llenyddiaeth ei chanfod drwy chwilio'n gynhwysfawr	3–5 mis	Adroddiad 15 tudalen
Adolygiad meta-ddadansoddol	Cyfuno canlyniadau cyfres o werthusiadau meintiol o ansawdd uchel	Amcangyfrif rhifiadol o effaith gyffredinol yr ymyriad a werthusir	3–5 mis	Adroddiad 15 tudalen
Adolygiad systematig llawn	Cael cyfosodiad tystiolaeth 'safon aur' o'r holl dystiolaeth sydd ar gael	Crynodeb cadarn iawn a systematig o'r dystiolaeth, gyda negeseuon allweddol ac argymhellion. Caiff llenyddiaeth ei chanfod drwy chwilio'n gynhwysfawr iawn	6–12 mis	Adroddiad 60 tudalen

Ffynhonnell: Y Sefydliad Cenedlaethol er Ymchwil i Addysg, *Investigating what works: through reviews*

Mae'n rhaid bod yn ofalus fan hyn. Mae cymeradwyo darparwyr masnachol fel Idox (neu'r peryg o'u henllibio os ydym yn feirniadol) yn diriogaeth anodd. Ond roeddem yn teimlo na allem osgoi crybwyll cyfunwyr tystiolaeth sy'n gofyn am ryw fath o ffi. A'r gwir amdani yw nad oes mynediad agored (hyd yma) i un o'r sianeli mwyaf ar gyfer ymchwil – cyfnodolion academiaidd – ac maent yn gofyn am ffi. Os ydym wedi methu unrhyw adnoddau cronfeydd data y credwch sy'n ddefnyddiol, **rhowch wybod i ni** er mwyn i ni allu eu hychwanegu at ein hadnoddau ar-lein.

## Cyngor arbenigol

Weithiau, ni fydd cronfeydd data goddefol yn ddigon. Efallai na fyddwn hyd yn oed yn siŵr o'n cwestiwn – nac felly pa ymchwil y dylem fod yn chwilio amdani. Dyma pryd gallai cael sgwrs gydag arbenigwr fod yn ddefnyddiol, felly da yw clywed am wasanaeth newydd sy'n cael ei ddatblygu yn yr Alban, **Evidence Bank**, a gaiff ei ddatblygu gan y Centre for Research on Families and Relationships. Yn ogystal â chynnig tystiolaeth, mae hefyd wedi peilota gwasanaeth cais arbennig. Datblygwyd y model mewn partneriaeth â sefydliadau a gwasanaethau yn y trydydd sector a'r sector cyhoeddus, ac mae'n cwmpasu plant, teuluoedd, perthnasau ac anabledd. Mae'r Evidence Bank yn cael ei ddatblygu ar hyn o bryd fel rhan o **What Works Scotland** a dylai fod ar-lein yn fuan.

Yn San Steffan a Whitehall, mae academyddion hefyd wedi dechau ar **Evidence Information Service**<sup>111</sup> ond megis dechrau mae hwn felly cadwch lygaid amdano pan fydd yn gwbl barod.

Os ydych chi mewn adran llywodraeth fawr yn San Steffan, yr Alban, Cymru neu Ogledd Iwerddon, rydych chi'n ddigon ffodus i gael dadansoddwyr, ymchwilwyr, ystadegwyr ac economegwyr mewnol i'w galw i'ch helpu i ganfod tystiolaeth a'i dehongli. Er enghraifft, mae'r **Government Social Research Service**.

Mae'r bennod hon wedi amlinellu amrywiaeth o adnoddau ymchwil hygyrch yn y DU ym maes polisi ac arfer cymdeithasol. Ond wedi ichi gasglu eich tystiolaeth, nid fyddwch eisiau iddi fod yn hel llwch. Mae angen i ni feddwl sut i'w chyfleu a gweithredu arni. Bydd y bennod nesaf yn canolbwyntio ar sut orau i wneud hynny.

### Negeseuon allweddol o Adran D

- Byddwch yn wiliadwrus o **chwiliadau hap a damwain ar-lein** – defnyddiwch ystorfeydd ymchwil dibynadwy.
- Manteisiwch ar y **crynodebau ymchwil gweithredol a hygyrch sydd ar gael am ddim**, megis y rhai sydd ar wefannau naw o ganolfannau What Works y DU, NHS Evidence, EPPI Centre a Campbell and Cochrane Collaborations.
- Darllenwch rai o'r **crynodebau bach am ddim** megis POSTNotes Senedd y DU neu Friffiau Tystiolaeth rheolaidd y Cyngor Ymchwil Economaidd a Chymdeithasol.
- Os na fedrwch ddod o hyd i'r hyn rydych chi'n chwilio amdano, dylech ystyried **comisiynu ymgynghoriaeth neu wasanaeth desg gymorth** i wneud crynodeb o ymchwil, megis asesiad tystiolaeth cyflym, adolygiad cwmpasu neu adolygiad systematig.
- Os ydych chi mewn adran llywodraeth neu awdurdod lleol, **defnyddiwch eich dadansoddwyr mewnol** i chwilio ar eich rhan.

# Adran E

## Sut dylech chi gyfleu eich canfyddiadau?

**Mae'r adran hon yn canolbwyntio ar sut i gyfleu eich neges unwaith y bydd gennyh chi'r dystiolaeth rydych chi ei hangen.**

Nid oes llawer o bwynt casglu corff cryf o dystiolaeth os yw'n syrthio ar glustiau byddar. Yr holl fwriad yw ei drosglwyddo i arfer a newid agweddau, credoau ac ymddygiad.

Mae cyfathrebu llwyddiannus yn golygu rhoi eich hun yn wirioneddol yn esgidiau'r gynulleidfa, gyda nifer o ganllawiau defnyddio ymchwil swyddogol yn dweud yn ddelfrydol y dylai hyn ddechrau gyda'r broses casglu tystiolaeth ei hun.<sup>112</sup>

Y peth gorau fyddai osgoi lledaenu'n oddefol, gan wthio gwybodaeth drwy'r drws yn y gobaith y bydd rhywfaint ohono'n glynu. Yn hytrach, mae angen i chi ymgysylltu'n weithredol â'ch cynulleidfaoedd.

Er enghraifft, mae gan ganllaw ymchwil<sup>113</sup> y Cyngor Ymchwil Feddygol rywfaint o gyngor defnyddiol ar ddefnyddio ymchwil ac mae'n cynnwys canllawiau ar sail ymchwil wedi'i chyhoeddi (yn hytrach na barn arbenigol). Mae'n ystyried polisi cymdeithasol yn ogystal â meddygaeth draddodiadol, ac mae'n cynnwys astudiaethau achos mewn meysydd megis atal trosedd ac iechyd cyhoeddus.

### Chwe awgrym o ganllaw'r Cyngor Ymchwil Feddygol ar gael gwrandawriad i'ch tystiolaeth

1. **Dylech gynnwys rhanddeiliaid** yn y broses o ddewis cwestiwn a dyluniad yr ymchwil er mwyn sicrhau perthnasedd.
2. Dylech ddarparu tystiolaeth **mewn ffordd integredig a graddol**: cynnwys adolygiadau, nid astudiaethau unigol, a chrynodebau o sawl hyd amrywiol sy'n caniatáu sganio'n gyflym.
3. **Dylech ystyried y cyd-destun**, a nodi'r elfennau perthnasol i wneud penderfyniadau, megis manteision, niwed a chostau.
4. **Dylech wneud argymhellion mor benodol** â phosibl.
5. Defnyddiwch **ymagwedd amlweddol sy'n cynnwys amrywiaeth o elfennau rhyngweithiol** – yn hytrach na didactig – o gyfarfodydd addysgol, adborth, pwyntiau atgoffa a phrosesau consensws lleol.
6. Mae rhoi ar waith yn llwyddiannus yn dibynnu ar newid ymddygiad – a hynny'n aml ymysg amrywiaeth eang o bobl. Mae hyn yn galw am **ddealltwriaeth wyddonol o'r ymddygiadau** sydd angen newid, y ffactorau sy'n cynnal yr ymddygiad presennol, ffactorau sy'n rhwystro a hwyluso newid, a'r arbenigedd i ddatblygu strategaethau i gyflawni newid yn seiliedig ar y ddealltwriaeth hon. Mae'n bosibl y bydd angen gwaith ymchwil pellach i helpu gyda'r broses o weithredu, a dylai timau ymchwil gweithredu gynnwys gwyddonydd ymddygiadol.

Ffynhonnell: Cyngor Ymchwil Feddygol (2013) *Developing and evaluating complex interventions: new guidance*

Yn anffodus, fel mae canllaw'r Cyngor Ymchwil Feddygol yn ei nodi, mae sylfaen y dystiolaeth ar gyfer gweithredu'n effeithiol yn parhau i fod yn gyfyngedig. Mae digon o 'ganllawiau arfer da' a chyngor, ond nid yw'n ymddangos bod y rhan fwyaf ohono'n seiliedig ar dystiolaeth – dim ond yr hyn sy'n cael ei gyfrif fel barn arbenigol. Ac mae'n ymddangos bod y dystiolaeth sydd gennym yn anghyson â nifer o arferion presennol.

Er enghraifft, mae'r obsesiwn â chrynodebau byr ar gyfer llunwyr polisi yn seiliedig ar dystiolaeth sigledig. Mewn hap-dreial wedi'i reoli amlweddol gan yr International Initiative for Impact Evaluation (3ie) ac eraill,<sup>114</sup> archwiliwyd effeithiolrwydd briff polisi o ran dylanwadu ar gredoau darllenwyr a'u hannog i weithredu.

*"Mae'r canlyniadau'n drawiadol ac nid ydynt yn tawelu meddwl y rhai hynny sydd, yn ôl 3ie, yn teimlo bod briffiau polisi'n bwysig,"* meddai Cyfarwyddwr Gweithredol 3ie ar y pryd, Howard White.<sup>115</sup> Er enghraifft, canfu'r treial nad yw newid natur y briff – megis ei hyd neu gynnwys safbwyntiau barn awdurdodol – yn amherthnasol i newid credoau. Fodd bynnag, roedd 'negesydd' y briff – hynny yw, pwy sy'n ei gyflwyno – o bwys i b'un a fyddai unrhyw un yn gweithredu arno. Mae tystiolaeth sylweddol o faes seicoleg gymdeithasol ar y 'negesydd' a byddwn yn rhoi sylw i hyn isod gan ei bod yn rhoi ychydig o gyfeiriad defnyddiol ar gyfleu polisi'n well.

## Canllawiau ar ddefnyddio ymchwil

- **Research Uptake Guidance** – Yr Adran dros Ddatblygu Rhyngwladol (y DU)
- **Communicating research for evidence-based policymaking: A practical guide for researchers in socio-economic sciences and humanities** – Y Comisiwn Ewropeaidd
- **Guide: Engaging with Policymakers** – National Coordinating Centre for Public Engagement (y DU)
- **Impact Toolkit** – Y Cyngor Ymchwil Economaidd a Chymdeithasol (y DU)
- **Helping researchers become policy entrepreneurs: How to develop engagement strategies for evidence-based policymaking** – Overseas Development Institute (y DU)

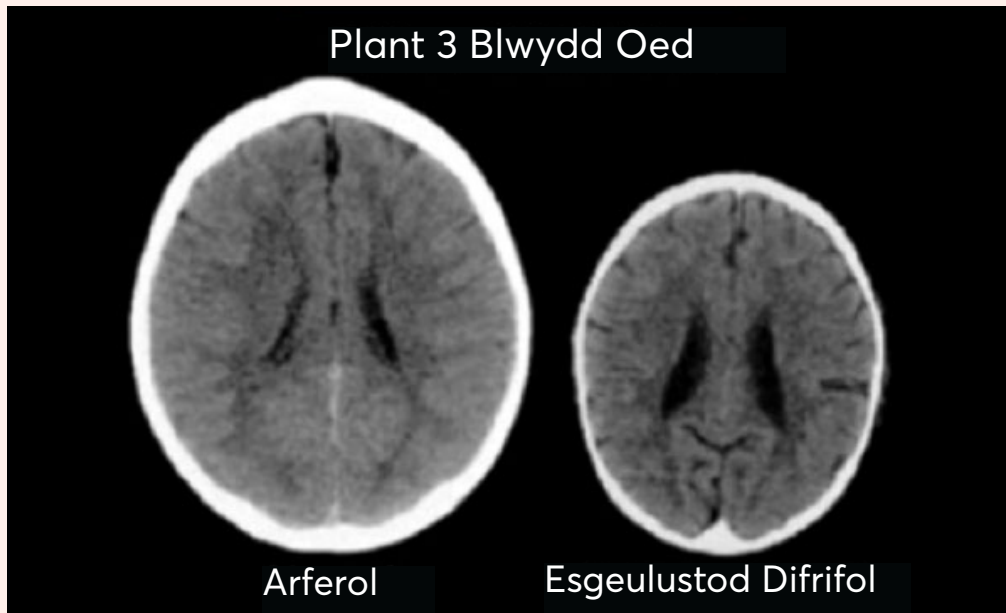
## Gwnewch hi'n hawdd, yn ddeniadol, yn gymdeithasol ac yn amserol

Os ydych chi am annog newid mewn ymddygiad yn seiliedig ar ymchwil, dylech ei gwneud hi'n Hawdd, yn Ddeniadol, yn Gymdeithasol ac yn Amserol (EAST – Easy, Attractive, Social, Timely). Mae'r pedair egwyddor syml hon yn seiliedig ar waith y Behavioural Insights Team ei hun a llenyddiaeth academiaidd ehangach:<sup>116</sup>

- **Hawdd** – e.e. **symleiddio'r neges**. Mwy na thebyg mai dyma un o'r negeseuon mwyaf cyffredin ar gyfer cyfleu ymchwil i gynulleidfaoedd lleyg. Gallai hyn fod mor syml hyd yn oed â fformat y briff ysgrifenedig. Mae ymchwil dilyn symudiad llygaid a gynhaliwyd ar gyfer y Post Brenhinol yn 2010<sup>117</sup> yn awgrymu bod **pobl fel arfer yn canolbwyntio ar benawdau, blychau a delweddau, tra bod testun manwl yn aml yn cael ei anwybyddu**. Fodd bynnag, gellir gor-bwysleisio'r arfer o symleiddio negeseuon. Mae rhywfaint o dystiolaeth bod cymhlethdod yn arwain at ganlyniadau dysgu mwy hir dymor, os ydych chi wir yn ceisio newid gwybodaeth a dealltwriaeth pobl.<sup>118</sup>
- **Deniadol** – e.e. **defnyddio delweddau neu bersonoli**. Mae hyn yn helpu i ddenu sylw (gweler y blwch ar y dudalen ganlynol ar defnyddio delweddau a sganiau MRI).
- **Cymdeithasol** – e.e. **defnyddio hanesion, a phobl go iawn**. Manteisiwch ar rwydweithiau er mwyn annog newid ymddygiad ehangach, ac annog pobl i wneud ymrwymiad i'w gilydd.
- **Amserol** – cawn ein dylanwadu'n fwy gan **gostau/manteision uniongyrchol** na rhai hwyrach (sy'n her i ymyriad cynnar) felly dylech ei gwneud yn gyfredol. Hefyd, dylech amseru'n iawn ac ysgogi pobl pan fyddan nhw fwyaf parod i dderbyn; gall gwneud yr un cynnig ar adegau gwahanol arwain at lefelau gwahanol iawn o lwyddiant.



## Pŵer delweddau – a sganiau'r ymennydd



Perry, B.D. (2002) Childhood experience and the expression of genetic potential: what childhood neglect tells us about nature and nurture. 'Brain and Mind'. 3: 79-100.

Bu'r ddelwedd hon yn ddylanwadol iawn ac fe'i rhoddwyd ar dudalen flaen adroddiad uchel ei barch gyda thystiolaeth dda i'w gefnogi ar gyfer y llywodraeth gan yr AS Graham Allen<sup>119</sup> a oedd yn dadlau'r achos dros ymyriad cynnar. Mae'r ymennydd ar yr ochr dde yn dangos effaith ar ymennydd plentyn sy'n dioddef o 'esgeulustod difrifol' a'r niwed am weddill ei fywyd, yn seiliedig o bosibl ar rianta gwael. Fodd bynnag, mae rhai wedi cwestiynu'r defnydd o ddelweddau MRI<sup>120</sup> o'r fath. Er enghraifft, beth oedd hanes achos y ddau blentyn uchod? Efallai bod gan y plentyn ar y dde anabledau difrifol eraill.

Ond gwyddom ba mor bwerys y gall niwro-ddelweddu fod wrth ddylanwadu ar lunwyr polisi – ac ar arbenigwyr eraill hyd yn oed. Canfu un hap-dreial wedi'i reoli fod cynnwys delweddau o ymennydd yn gysylltiedig â chanfyddiad uwch o deilyngdod gwyddonol. Dangoswyd papurau academiaidd, a oedd union yr un fath, i fyfyrwyr prifysgol, a'r unig wahaniaeth oedd bod un yn cynnwys delweddau o ymennydd, ac nid oedd delweddau o ymennydd ar y llall (ac roedd un yn cynnwys delweddau eraill, megis graffiau bar traddodiadol). Credai'r bobl fod i'r papur â delwedd o ymennydd lawer mwy o hygrededd gwyddonol, heb sylweddoli mai union yr un testun ac ymchwil oedd iddynt.<sup>121</sup>

### Negeseuon allweddol o Adran E

- Mae cyfathrebu'n llwyddiannus yn golygu **dangos empathi tuag at eich cynulleidfa darged** – yn ddelfrydol o ddechrau'r broses o gasglu tystiolaeth.
- Dylech osgoi lledaenu goddefol. Yn lle hynny, mae angen i chi **ymgysylltu'n weithredol â'ch cynulleidfaoedd**.
- Os oes yn rhaid i chi fod yn fwy goddefol wrth gyfathrebu, defnyddiwch awgrymiadau'r Behavioural Insights Team ar newid ymddygiad. Dylech ei **Gwneud hi'n Hawdd, yn Ddeniadol, yn Gymdeithasol ac yn Amserol**.

# Crynodeb

P'un a ydych chi'n lluniwr polisi neu'n arweinydd mewn elusen neu awdurdod lleol, mae'n annhebygol y bydd tystiolaeth yn 'cydweddu' yn dda â'r holl ffynonellau eraill o wybodaeth sydd ar gael. Rhan o'r gymysgedd yn unig yw tystiolaeth ymchwil, ac felly y dylai fod.

Mae sut i fynd ati i wneud penderfyniadau clyfrach yn gyffredinol yn ganllaw arfer newydd ynddo'i hun. Awgrymiadau a rhestrau gwirio yn unig y gellir eu rhoi yma. Ni fydd byth fformiwla hud ar sut i 'wneud' polisi sy'n seiliedig ar dystiolaeth. Fodd bynnag, mae modelau a theclynnau ar gael i helpu i wneud penderfyniadau. Ar gyfer llunwyr polisi, mae tîm **Evidence for Policy Design Prifysgol Harvard** wedi creu teclynnau ar-lein megis 'coed-penderfyniad' neu 'fatricks dadansoddi polisi' i helpu i bwysu a mesur y darnau cystadleuol o wybodaeth sydd y tu ôl i bolisi, yn sgil arian gan DFID y DU.<sup>122</sup> Mae'r teclynnau gwneud penderfyniadau hyn yn ddefnyddiol ond ni fyddan nhw byth yn disodli barn broffesiynol ym myd cymhleth a blêr polisi cymdeithasol. Gallant fod yn offeryn neu'n rhestr wirio i'w gefnogi. Yn hytrach na gwneud ein rhestr ein hunain, mae gennym ddau brif argymhelliad ar gyfer defnyddio tystiolaeth yn sail i wneud penderfyniadau gwell:

## 1. Rhoi newidiadau neu gyngor sy'n seiliedig ar dystiolaeth ar waith sydd mor gryf â phosibl (ac i'r gwrthwyneb, bod yn wylidwrus o newidiadau neu gyngor os yw'r dystiolaeth yn wan)

Cyngor y canllaw yw po gryfaf yw'r dystiolaeth, yn enwedig os yw'n seiliedig ar fwy nag un efelychiad ac adolygiadau systematig yn hytrach nag un astudiaeth yn unig, y mwyaf hyderus y dylech chi fod o'i gynnwys yn rhan o'ch penderfyniad.

Yr hyn na ddylem ei wneud yw gwneud awgrymiadau polisi mwy beiddgar yn seiliedig ar dystiolaeth wannach. Efallai bod hynny'n ymddangos yn amlwg i chi, ond nid yw'n ymddangos bod hynny'n digwydd ym maes llunio polisi. Canfu adolygiad y Swyddfa Archwilio Genedlaethol ar dystiolaeth llywodraeth (gweler tudalen 29) fod y gwerthusiadau gwannaf yn fwy tebygol o wneud honiadau beiddgar heb gafeat am effeithiau cadarnhaol y polisi a archwiliwyd.<sup>123</sup> Os oes tystiolaeth wan, neu nad oes tystiolaeth yn bodoli, byddwch yn onest am hynny gan ddweud hynny. Gêm beryglus yw honni gormod. Ac er y gwyddom mor rhwystredig yw gweld casgliad yn dweud bod 'angen mwy o ymchwil' mewn adroddiadau neu gyngor polisi, y mae hefyd yn wir: ni ddaw'r angen i ddeall ac ymchwilio byth i ben.

## 2. Byddwch yn ymwybodol o dystiolaeth wrth wneud penderfyniadau

Yr ail ddarn o gyngor yw bod yn hunanymwybodol am dueddiadau gwybyddol posibl wrth i chi wneud penderfyniadau. Mae bod yn ymwybodol o'r ffaith y gallech fod yn neidio i'r casgliadau anghywir – hyd yn oed pan fyddwch chi wedi defnyddio'r ymchwil gorau posibl – yn hanfodol.

Yn olaf, dymunwn bwysleisio mai'r neges bwysicaf un yn y canllaw hwn yw priodolrwydd: 'llunio'r gwadn fel bo'r droed'. Mae angen ichi feddwl am y math cywir o ymchwil i gyd-fynd â'ch anghenion.

Mae'n rhaid iddo fod yn gymesur – byddai hap-dreial wedi'i reoli sy'n enfawr ac amlweddol, yn wallgof ar gyfer menter gymdeithasol fach yng nghefn gwlad. Ond yn yr un modd, dylai rhaglen gymdeithasol £50 miliwn sydd i bara am ddegawd wirioneddol fuddsoddi mewn tystiolaeth.

Mae'n rhaid i'r math o dystiolaeth hefyd gyd-fynd â'ch cwestiwn. A yw'n cydweddu â lle'r ydych o ran tyfu rhaglen, polisi neu arfer? I'ch helpu i lunio barn ar hynny, defnyddiwch rai o'r fframweithiau ffurfiol, fel Safonau Tystiolaeth Nesta, os yw eich cwestiwn yn ymwneud ag effaith. Hefyd, ystyriwch ble'r ydych yn nhroell Nesta – ai dysgu gan eraill fyddai eich dewis gorau, neu a yw'n bryd arbrofi'n ffurfiol â'ch syniad sy'n tyfu?

A pheidiwch â chael eich dychryn gan gost – rydym wedi rhoi rhai cyfeiriadau at ble i ddod o hyd i ymchwil sydd eisoes ar gael, a hynny am ddim. Mae'n dod yn haws cael gafael ar ymchwil. Mae academyddion dan fwy o bwysau i gamu y tu allan i waliau eu prifysgol a'ch helpu ym maes llunio polisi, gwaith elusennol neu gyflawni gwasanaethau lleol da.

Mae hefyd angen inni fod yn fwy clyfar o ran sut i gyfleu tystiolaeth. Prin iawn y bydd tystiolaeth yn siarad ar ei rhan ei hun. Neges gyfathrebu gyffredin yw symleiddio, darlunio, dangos empathi a bod yn 'gyfieithiadol' – hynny yw, rhyngweithio gyda'ch cynulleidfa. Yn y misoedd a'r blynyddoedd sydd i ddod, byddwn yn gwneud hynny'n union: creu fideos a chyfryngau digidol rhyngweithiol ar gyfer dysgu cyfunol, ac efallai, annwyl ddarllenwr, eich cyfarfod wyneb yn wyneb yn un o'n Dosbarthiadau Meistr Tystiolaeth parhaus.

# Atodiad 1

## Sail resymegol i dystiolaeth a mathau o dystiolaeth sy'n ofynnol ar gyfer llunwyr polisi

Ffynhonnell: Louise Shaxson (2014) *Investing in Evidence; Lessons from the UK Department for Environment, Food and Rural Affairs*.

Pennawd	'Y cwestiynau mawr'	Sail resymegol i anghenion tystiolaeth	Mathau o dystiolaeth sy'n ofynnol
A: <b>Deall y cyd-destun;</b> prosesau a ffenomenau hanfodol, sylfeini a meincnodau	Ble rydyn ni nawr?	<ul style="list-style-type: none"> <li>Casglu a dadansoddi data sydd ar gael / data newydd</li> <li>Gwerthuso risgiau, materion ac ansicrwydd</li> </ul>	<ul style="list-style-type: none"> <li>Adolygiadau o wybodaeth gyfredol</li> <li>Arolygon o ddata cymdeithasol ac amgylcheddol</li> <li>Ymchwil ar achosiaeth</li> <li>Asesiad risg</li> </ul>
B: <b>Datblygu</b> modelau, methodolegau ac offer	Ble rydyn ni'n mynd?	<ul style="list-style-type: none"> <li>Deall ysgogwyr a thuaddiadau cyfredol</li> <li>Rhagweld ysgogwyr a thuaddiadau'r dyfodol</li> <li>Asesu goblygiadau ar gyfer canlyniadau polisi</li> </ul>	<ul style="list-style-type: none"> <li>Dadansoddi sensitifrwydd</li> <li>Sganio'r gorwel</li> <li>Rhagweld sefyllfaoedd</li> <li>Modelu effeithiau a chanlyniadau</li> </ul>
C: <b>Datblygu a defnyddio'r</b> sylfaen dystiolaeth i helpu i <b>osod targedau a ffurfio polisi</b>	Lle rydyn ni eisiau bod dros y 5-10 mlynedd nesaf?	<ul style="list-style-type: none"> <li>Deall gwerth economaidd/ cymdeithasol newid</li> <li>Deall ymarferoldeb/cost newid</li> <li>Cyd-drafod nodau</li> </ul>	<ul style="list-style-type: none"> <li>Ymchwil economaidd a chymdeithasol</li> <li>Prosesau ymgysylltu ymgynghori</li> <li>Astudiaethau peilot ac ymarferoldeb</li> <li>Arolygon o'r farchnad</li> </ul>
D: <b>Datblygu a gwerthuso opsiynau/datrysiadau</b>  E: Penderfyniadau gorau posibl a gweithredu effeithiol drwy <b>gyfathrebu</b> , ymgysylltu ac ymgynghori i ddylanwadu ar newid	Sut fyddwn ni'n cyrraedd yno?	<ul style="list-style-type: none"> <li>Nodi/gwerthuso opsiynau cyfredol</li> <li>Nodi/datblygu datrysiadau newydd</li> <li>Gwerthuso opsiynau newydd/ hen opsiynau</li> </ul>	<ul style="list-style-type: none"> <li>Astudiaethau opsiwn/ gwerthuso</li> <li>Asesiadau effaith rheoleiddiol</li> <li>Ymyriadau i hybu arloesedd</li> </ul>
F: <b>Monitro cynnydd</b> tuag at bolisi/targedau rhaglen  G: <b>Gwerthusiad</b> polisi/ rhaglen	Pa mor dda wnaethon ni?	<ul style="list-style-type: none"> <li>Monitro cynnydd</li> <li>Gwerthuso polisiau a rhaglenni</li> <li>Dysgu gwersi</li> </ul>	<ul style="list-style-type: none"> <li>Gwerthusiadau rhyngddisgyblaethol</li> <li>Prosesau gwerthuso ymgynghori</li> </ul>

## Atodiad 2

# Dyluniadau arbrolol ar gyfer gwerthuso ymyriadau cymhleth

Daw'r rhestr hon o bum dyluniad arbrolol gwahanol o ddogfen *Developing and evaluating complex interventions: new guidance*, a baratowyd ar ran y Cyngor Ymchwil Feddygol (2013).

### 1. Hap-dreialon unigol

Caiff unigolion eu nodi ar hap naill ai i dderbyn ymyriad arbrolol, neu ddewis amgen megis triniaeth safonol, plasebo neu barhau ar restr aros. Caiff treialon o'r fath weithiau eu hystyried yn amherthnasol i ymyriadau cymhleth, ond mae nifer o amrywiadau ar y dull sylfaenol, ac yn aml gellir canfod datrysiadau i'r problemau technegol a moesegol sy'n gysylltiedig â dewis ar hap.

### 2. Hap-dreialon clwstwr

Yn aml, ystyrir bod achos o halogi'r grŵp rheoli, gan arwain at amcangyfrifon â thuedd o faint yr effaith, yn anfantais i hap-dreialon ymyriadau ar lefel poblogaeth, ond mae hap-dreialon clwstwr, a ddefnyddir yng ngwaith ymchwil y gwasanaeth iechyd, yn un ateb. Yma, mae grwpiau megis cleifion mewn meddygfa neu denantiaid mewn cynllun tai yn cael eu dewis ar hap i dderbyn yr ymyriad arbrolol neu'r ymyriad rheoli.

### 3. Dyluniadau lletem risiog

Gellir defnyddio dyluniad lletem risiog i oresgyn gwrthwynebiadau ymarferol neu foesegol i werthuso ymyriad yn arbrolol lle mae rhywfaint o dystiolaeth o effeithiolrwydd ar ei gyfer, neu lle nad yw'n bosibl i'r boblogaeth gyfan fanteisio arno ar unwaith. Mae'n caniatáu i hap-dreial wedi'i reoli gael ei gynnal heb achosi oedi o ran rhoi'r ymyriad ar waith yn gyffredinol. Yn y pen draw, mae'r boblogaeth gyfan yn derbyn yr ymyriad, ond gyda dewis ar hap yn rhan o roi'r ymyriad ar waith yn raddol.

### 4. Treialon hoff ddewis a dyluniadau hap-ganiatâd

Yn aml gellir goresgyn rhwystrau ymarferol neu foesegol i ddewis ar hap drwy ddefnyddio dyluniadau nad ydynt yn safonol. Lle mae gan gleifion hoff ddewis ymysg triniaethau, gallai fod yn briodol dyrannu triniaeth yn seiliedig ar hoff ddewis y cleifion, neu ddewis cleifion ar hap cyn cael caniatâd.

### 5. Dyluniadau N-o-1

Nod treialon confensiynol yw amcangyfrif effaith gyfartalog ymyriad ar boblogaeth, a rhoi ychydig wybodaeth am amrywiant person mewn ymateb i ymyriadau, neu am y mecanweithiau mae ymyriadau effeithiol yn eu defnyddio i gyflawni newid. Gellir defnyddio treialon N-o-1, lle mae unigolion yn derbyn ymyriadau gyda'r drefn neu'r amserlen wedi'i dewis ar hap, i asesu newid rhwng pobl a newid oddi mewn i bobl, ac ymchwilio i gyfryngwyr y newid hwnnw a ragwelwyd yn ddamcaniaethol.

# Atodiad 3

## Canllawiau ar ansawdd tystiolaeth, dyluniadau a dulliau

Social Researchers' Method Guides, Llywodraeth yr Alban – cyflwyniadau byr i rai o'r dulliau mwyaf cyffredin a ddefnyddir mewn ymchwil gwyddor gymdeithasol

[www.gov.scot/Topics/Research/About/Social-Research/Methods-Guides](http://www.gov.scot/Topics/Research/About/Social-Research/Methods-Guides)

The Bond Evidence Principles a'r rhestr wirio - wedi'u dylunio'n benodol ar gyfer Sefydliadau Anllywodraethol

[www.bond.org.uk/effectiveness/monitoring-and-evaluation](http://www.bond.org.uk/effectiveness/monitoring-and-evaluation)

Evidence for Success: the guide to getting evidence and using it – canllaw ymarferol ar gyfer sefydliadau'r trydydd sector gan Evaluation Support Scotland a KTN

[www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised\\_april\\_2015-web.pdf](http://www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised_april_2015-web.pdf)

Research Guide for Third Sector Organisations

[www.ncb.org.uk/media/858179/research\\_guide\\_for\\_third\\_sector\\_organisations.pdf](http://www.ncb.org.uk/media/858179/research_guide_for_third_sector_organisations.pdf)

Developing and evaluating complex interventions: new guidance, y Cyngor Ymchwil Feddygol – yn darparu canllawiau ar ddatblygu, gwerthuso a rhoi ymyriadau cymhleth ar waith i wella iechyd

[www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/](http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/)

Swyddfa'r Llywodraeth ar gyfer Gwyddoniaeth – Canllawiau Prif Ymgynghorydd Gwyddonol y Llywodraeth ar Ddefnyddio Cyngor Peirianeg a Gwyddonol wrth Lunio Polisi

[www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf)

Y Sefydliad Cenedlaethol dros Ragoriaeth mewn Iechyd a Gofal (NICE) – The Guidelines Manual: Reviewing the Evidence

[www.nice.org.uk/article/pmg6/chapter/6-reviewing-the-evidence](http://www.nice.org.uk/article/pmg6/chapter/6-reviewing-the-evidence)

DFID – How to Note: Assessing the Strength of Evidence

[www.gov.uk/government/publications/how-to-note-assessing-the-strength-of-evidence](http://www.gov.uk/government/publications/how-to-note-assessing-the-strength-of-evidence)

Trysorlys Ei Mawrhydi – The Magenta Book: Guidance for Evaluation

[www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/220542/magenta\\_book\\_combined.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf)

Civil Service Learning – Policy Profession: Skills and Knowledge Framework

[civilservicelearning.civilservice.gov.uk/sites/default/files/policy\\_profession\\_skills\\_and\\_knowledge\\_framework\\_jan2013web.pdf](http://civilservicelearning.civilservice.gov.uk/sites/default/files/policy_profession_skills_and_knowledge_framework_jan2013web.pdf)

Trysorlys Ei Mawrhydi, DECC a DEFRA – Quality in policy impact evaluation; deall effeithiau polisi o ddylanwadau eraill (Atodiad i'r Magenta Guide)

[www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/190984/Magenta\\_Book\\_quality\\_in\\_policy\\_impact\\_evaluation\\_QPIE\\_.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation_QPIE_.pdf)

Cynghorau Ymchwil, Prifysgolion y DU et al. – Concordat y DU i gefnogi gonestrwydd ymchwil

[www.universitiesuk.ac.uk/highereducation/Pages/Theconcordattosupportresearchintegrity.aspx#.Vd2X5fIVhBc](http://www.universitiesuk.ac.uk/highereducation/Pages/Theconcordattosupportresearchintegrity.aspx#.Vd2X5fIVhBc)

### Canllawiau ar gomisiynu ymchwil

Social Research Association – Commissioning Social Research good practice guide

[the-sra.org.uk/sra\\_resources/research-commissioning/](http://the-sra.org.uk/sra_resources/research-commissioning/)

Market Research Society guide to 'buying research'

[www.mrs.org.uk/intelligence/research\\_buyers\\_guide](http://www.mrs.org.uk/intelligence/research_buyers_guide)

# Ôl-nodiadau

- Sefydliad yr Heddlu (2013) Policing and Crime Reduction: The evidence and its implications for practice. Gweler tudalen 11: [http://www.police-foundation.org.uk/uploads/holding/projects/policing\\_and\\_crime\\_reduction.pdf](http://www.police-foundation.org.uk/uploads/holding/projects/policing_and_crime_reduction.pdf)
- Sense about Science (2015) Making Sense of Crime: Evidence on the causes of crime and policies to reduce it. Gweler tudalen 17.
- OECD (2015) Students, Computers and Learning: Making the Connection. Gweler: <http://www.oecd.org/publications/students-computers-and-learning-9789264239555-en.htm>
- Am drosolwg o dystiolaeth ar 'beth sy'n gweithio' yn yr ystafell ddosbarth ar gyfer cyrhaeddiad disgyblion, gweler Pecyn Cymorth y Sefydliad Gwaddol Addysgol: <https://educationendowmentfoundation.org.uk/toolkit/toolkit-a-z/>
- Daw hwn o'r 'Policy Profession Skills and Knowledge Framework' (gweler ôl-nodyn 9) ond mae'r un mor berthnasol i'n cynulleidfa yn y canllaw hwn, sy'n gweithio y tu allan i bolisi, mewn awdurdodau lleol ac elusennau.
- Sackett, D., Rosenberg, W., Gray, M., Haynes, B. a Richardson, S. (1996) Evidence based medicine: what it is and what it isn't. 'British Medical Journal'. 312:71. Gweler: <http://www.bmj.com/content/312/7023/71>
- Daeth y diffiniad mewn cyfnod pan roedd y rhan fwyaf o benderfyniadau meddygol yn cael eu gwneud yn seiliedig ar brofiad, awdurdod a bri. Nid oedd arfer feddygol yn seiliedig ar y dystiolaeth wyddonol orau posibl. Mae sylwebwyr ac ymchwilwyr fel Ben Goldacre wedi dadlau bod polisi cymdeithasol yn yr un man â meddygaeth 20 mlynedd yn ôl: awdurdod sy'n teyrnasu o hyd, nid ymchwil.
- Graber, M.L., Franklin, N. a Gordon, R. (2005) Diagnostic Error in Internal Medicine. 'Internal Medicine.' 165(13). 1493- 1499. Gweler: <http://archinte.jamanetwork.com/article.aspx?articleid=486642&res ultclick=1>
- Civil Service Learning (2013) 'Policy Profession Skills and Knowledge Framework.' London: Civil Service Learning. Gweler: [https://civilservicelearning.civilservice.gov.uk/sites/default/files/policy\\_profession\\_skills\\_and\\_knowledge\\_framework\\_jan2013web.pdf](https://civilservicelearning.civilservice.gov.uk/sites/default/files/policy_profession_skills_and_knowledge_framework_jan2013web.pdf)
- <http://www.oxforddictionaries.com/definition/english/evidence>
- Frost, S. et al. (2006) 'The Evidence Guide; Using Research and Evaluation in Social Care and Allied Professions.' Llundain: Barnardo's
- Nutley, S., Powell, A. a Davies, H. (2013) 'What Counts as Good Evidence?' Llundain: Y Gynghrair ar gyfer Tystiolaeth Ddefnyddiol
- Addaswyd o Shaw, I. (1999) 'Qualitative Evaluation.' Llundain: Sage.
- Tetlock, P.E. (2005) Expert Political Judgement: How Good Is It? How Can We Know?. New Jersey: Princeton University Press
- Vermeren, P. (2015) 'Why the CEBMa model is misleading.' Gweler: [http://evidencebasedhrm.be/wp-content/uploads/2015/08/Why-the-Cebma-model-is-misleading\\_august\\_2015.pdf](http://evidencebasedhrm.be/wp-content/uploads/2015/08/Why-the-Cebma-model-is-misleading_august_2015.pdf)
- Ariely, D. (2009) Predictably Irrational: The Hidden Forces that Shape Our Decisions. Llundain: HarperCollins.
- Gweler hefyd White, H. a Phillip, D. (2012) 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework' – papur gweithredol gan yr International Initiative for Impact Evaluation sy'n trafod sut gall gwerthuswyr eu hunain fod yn dueddol iawn. "Bias affects all evaluation approaches and methods; it affects the choices that evaluators make, the questions that they ask, the data they collect and the way that they carry out analysis" (tudalen 24).
- Caiff y posibilrwydd bod gweithwyr cymdeithasol yn cael pethau'n anghywir ei gymhlethu gan faich trwm y penderfyniadau a wrânt – mae'n rhaid iddynt wneud penderfyniadau dilynol anodd yn ystod un diwrnod, sy'n achosi perygl o 'flinder penderfynu'. I weld trafodaeth ar yr ymchwil hon gweler tudalen 7 yn Breckon, J. a Hay, J. (2015) 'Knowing How to Protect; Using Research Evidence to Protect Children from Harm'. Llundain: Y Gynghrair ar gyfer Tystiolaeth Ddefnyddiol / NSPCC. Gweler: <http://www.alliance4usefulevidence.org/assets/NSPCC-report-FINAL-V-for-printers.pdf>
- Kassin, S.M., Dror, I.E. a Kukuckaa, J. (2013) The forensic confirmation bias: Problems, perspectives, and proposed solutions. 'Journal of Applied Research in Memory and Cognition.' 2(1). 42-52. Gweler: [http://cognitiveconsultantsinternational.com/Dror\\_JARMAC\\_forensic\\_bias.pdf](http://cognitiveconsultantsinternational.com/Dror_JARMAC_forensic_bias.pdf)
- Kahneman, D., Lovallo, D. a Sibony, O. (2012) 'The Big Idea: Before You Make That Big Decision...'. Gweler: <https://hbr.org/2011/06/the-big-idea-before-you-make-that-big-decision>
- Flyvbjerg, B., Skamris Holm, M.K. a Buhl, S.L. (2002) Underestimating Costs in Public Works Projects – Error or Lie. 'Journal of the American Planning Association.' 68(3). 279-295.
- Gweler tudalen 7 yn Lewis, P.G. (2013) 'Policy Thinking, Fast and Slow: A Social Intuitionist Perspective on Public Policy Processes'. Papur Cyfarfod Blynyddol APSA 2013, Cyfarfod Blynyddol 2013 American Political Science Association. Gweler: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2300479](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2300479)
- Gweler Cairney, P. (2014) 'The Psychology of Evidence Based Policymaking: Who Will Speak For the Evidence if it Doesn't Speak for Itself?' Gweler: <https://paulcairney.wordpress.com/2014/04/07/1288/>
- Kahneman D. a Klein G. (2009) Conditions for intuitive expertise: a failure to disagree. 'American Psychology.' 64(6). 515-26. Gweler: <http://www.ncbi.nlm.nih.gov/pubmed/19739881>
- Gornall, J. (2008) Does cot death still exist? 'British Medical Journal.' 336(7639). 302-304. Gweler: <http://dx.doi.org/10.1136/bmj.39455.496146.AD>
- Nesta (2013) 'Understand how innovation works.' Fideo ar gael yn <http://www.nesta.org.uk/resources/understand-how-innovation-works>
- Gabriel, M. (2014) 'Making it Big: Strategies for scaling social innovations.' Llundain: Nesta. Gweler: <http://www.nesta.org.uk/publications/making-it-big-strategies-scaling-social-innovations>
- White, H. (2015) 'Learning what works and why from evidence of what doesn't work.' Gweler: <http://www.alliance4usefulevidence.org/learning-what-works-and-why-from-evidence-of-what-doesnt-work/>
- Nesta a TSIP (2014) 'Guidance for Developing a Theory of Change for Your Programme.' Gweler: [https://www.nesta.org.uk/sites/default/files/theory\\_of\\_change\\_guidance\\_for\\_applicants.pdf](https://www.nesta.org.uk/sites/default/files/theory_of_change_guidance_for_applicants.pdf)
- Gweler tudalen 3 yn Stern, E. (2015) 'Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.' Llundain: Y Gronfa Loteri Fawr, Comic Relief a'r Adran dros Ddatblygu Rhyngwladol.
- Johnson, S. (2011) 'Impact of social science on policy: Sure Start case study.' Adroddiad i'r Cyngor Ymchwil Economaidd a Chymdeithasol (ESRC)
- Mulgan, G. (2015) 'The six Ws: a formula for what works.' Llundain: Nesta. Gweler: <http://www.nesta.org.uk/blog/six-ws-formula-what-works>
- Nesta (2013) 'What is ethnographic research?' Fideo ar gael yn <https://vimeo.com/77509484>
- Petticrew, M. a Roberts, H. (2003) Evidence, hierarchies and typologies: horses for courses. 'Journal of Epidemiology and Community Health.' 57. 527-529.
- Am restr awdurdodol a chynhwysfawr o fframweithiau a dulliau ymchwil gwyddor gymdeithasol gweler Luff, R., Byatt, D. a Martin, D. (2015) 'Review of the Typology of Research Methods within the Social Sciences.' Llundain: ESRC/National Centre for Research Methods. Y deipoleg hon yw un o'r rhannau sydd wedi'i lawrlwytho fwyaf ar wefan NCRM. Gweler: <http://eprints.ncrm.ac.uk/3721/>
- Gwnaed yr un pwynt mewn canllaw diweddar yn 2015 ar werthuso effaith. Yn ôl papur gan Elliot Stern, a gomisiynwyd gan y Gronfa Loteri Fawr, Bond, Comic Relief a'r Adran dros Ddatblygu Rhyngwladol, un o'r problemau sy'n wynebu'r rhai sy'n mynd i'r afael â gwerthusiadau effaith yw: "it is often presented as a technical or methodological question only accessible to experts or researchers. To some extent this is true but the main arguments, logics and choice-points are more accessible."



37. Yr Adran dros Ddatblygu Rhyngwladol (2014) 'How to Note: Assessing the Strength of Evidence.' Gweler: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/291982/HTN-strength-evidence-march2014.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf)
38. Am fwy o wybodaeth am bwysigrwydd hanesion, gweler Greenhalgh, T. (2014) 'Stories or numbers OR stories and numbers.' Gweler: <http://www.alliance4usefulevidence.org/stories-or-numbers-or-stories-and-numbers/>
39. Am feirniadaeth sylfaenol gweler Frideres, J. S. (1992) 'Participatory research: An illusionary perspective.' in Frideres, J.S. (ed.) A world of communities: Participatory research perspectives. North York, Ontario: Captus University Publications. Mae Frideres, wrth gwrs, wedi'i feirniadu'n fawr gan y gymuned ymchwil gyfranogol.
40. Mabwysiadwyd hwn o Dabl 2, tudalen 18 yn Stern, E. (2015) 'Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.'
41. White, H. a Phillips, D. (2012), 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework.' Gweler: [http://www.3ieimpact.org/media/filer\\_public/2012/06/29/working\\_paper\\_15.pdf](http://www.3ieimpact.org/media/filer_public/2012/06/29/working_paper_15.pdf)
42. <http://www.nesta.org.uk/publications/better-public-services-through-experimental-government>
43. Blair, P.S., Fleming, P.J., Smith, I.J., Platt, M.W., Young, J., Nadin, P. et al. (1999) Babies sleeping with parents: case- control study of factors influencing the risk of the sudden infant death syndrome. 'British Medical Journal.' 319 (7223). 1457-1462.
44. Y Cyngor Ymchwil Feddygol (2013) 'Developing and evaluating complex interventions: new guidance.' Gweler: <http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>
45. Gweler: <http://www.alliance4usefulevidence.org/publication/spark-guide-to-testing-social-policy-innovation/>
46. Gweler tudalen 5 yn Stern, E. (2015) 'Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations in the Voluntary and Community Sector.'
47. Gweler tudalen 21, ibid.
48. Carter, R. (2012) 'Helpdesk Research Report: Theory-based evaluation approach.' Birmingham: Governance and Social Development Resource Centre, University of Birmingham. Gweler: <http://www.gsdr.org/docs/open/hdq872.pdf>
49. Westthorp, G. (2014) 'Realist Evaluation: An Introduction.' Llundain: Methods Lab, Overseas Development Institute. Gweler: <http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf>
50. Mayne, J. (2012) Contribution Analysis: Coming of Age?. 'Evaluation.' 18(3). 270-280.
51. Am ddadansoddiad rhwydwaith ac olrhain proses gweler: Befani, B. a Mayne, J. (2014) Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation. 'IDS Bulletin.' 45(6). 17-36. Gweler: <http://onlinelibrary.wiley.com/doi/10.1111/1759-5436.12110/abstract>
52. Baptist, C. a Befani, B. (2015) 'Coffey How To Guide: Qualitative Comparative Analysis – A Rigorous Qualitative Method for Assessing Impact.' Gweler: <http://www.coffey.com/assets/Ingenuity/Qualitative-Comparative-Analysis-June-2015.pdf>
53. Gweler Social Researchers' Method Guide Llywodraeth yr Alban ar Ddadansoddi Cyfraniad: <http://www.gov.scot/resource/doc/175356/0116687.pdf>
54. Er enghraifft, yr hen ganllaw gan y diweddar arbenigwr gwerthuso gwych Weiss: Weiss, C. (1998) Evaluation: methods for studying programs and policies. Efrog Newydd: Prentice Hall.
55. <http://www.thinknpc.org/blog/getting-started-with-cost-benefit-analysis/>
56. Swyddfa'r Cabinet (2007) 'A guide to Social Return on Investment.' Gweler: [https://www.bond.org.uk/data/files/Cabinet\\_office\\_A\\_guide\\_to\\_Social\\_Return\\_on\\_Investment.pdf](https://www.bond.org.uk/data/files/Cabinet_office_A_guide_to_Social_Return_on_Investment.pdf)
57. Trysorlys Ei Mawrhydi (2011) 'The Green Book: Appraisal and Evaluation in Central Government.' Llundain: Trysorlys Ei Mawrhydi.
58. <https://www.nao.org.uk/report/choosing-the-right-fabric-3/>
59. <https://www.gov.uk/government/publications/supporting-public-service-transformation-cost-benefit-analysis-guidance-for-local-partnerships>
60. Petticrew, M. a Roberts, H. (2003) Evidence, hierarchies and typologies: horses for courses. 'Journal of Epidemiology and Community Health.' 57. 527-529.
61. Ioannidis, J.A. (2005) Contraindications and Initially Stronger Effects in Highly Cited Clinical Research. 'The Journal of the American Medical Association.' 294(2). 218-228.
62. Wedi'i ddyfynnu yn The Economist (2013) 'Unreliable research; trouble in the lab.' 409(8858). 24.
63. Easterbrook, P.J., Gopalan, R., Berlin, J.A. a Matthews, D.R. (1991) Publication bias in clinical research. 'The Lancet.' 337(8746). 867-872.
64. Mae'r gwyddorau cymdeithasol yn dioddef o duedd cyhoeddi difrifol; canfu'r arwlwg mai anaml iawn mae 'canlyniadau nwl' yn gweld golau dydd.
65. Wolford, B. (2014) 'Big Pharma Plays Hide-the-Ball With Data.' Gweler: <http://europe.newsweek.com/medical-science-has-data-problem-284066?rm=eu>
66. Sense about Science (2006) 'I don't know what to believe: Making sense of science stories.' Gweler: <http://www.senseaboutscience.org/resources.php/16/i-dont-know-what-to-believe>
67. Gweler tudalen 10 yn nogfen Swyddfa'r Llywodraeth ar gyfer Gwyddoniaeth (2010) 'The Government Chief Scientific Adviser's Guidelines on the Use of Scientific and Engineering Advice in Policy Making.' Gweler: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf)
68. Gweler tudalen 14 yn [http://www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised-april\\_2015-web.pdf](http://www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised-april_2015-web.pdf)
69. Y Swyddfa Archwilio Genedlaethol (2013) 'Evaluation in Government.' Gweler: <http://www.nao.org.uk/report/evaluation-government/#>
70. Gweler er enghraifft, y canllaw hwn fel atodiad i'r Magenta Guide: Trysorlys Ei Mawrhydi, DECC a DEFRA (2012) 'Quality in policy impact evaluation; understanding the effects of policy from other influences.' Mae'r canllaw yn dangos bod: 'higher quality research designs can help meet the challenge of attributing measured outcomes to the policy in question (as opposed to other influences), whereas lower quality designs reduce confidence in whether it was the policy that achieved those outcomes' (tudalen 5).
71. Am restr a disgrifiad byr o'r graddfeydd a'r hierarchaethau hyn, gweler Atodiad 1 yn Nutley, S., Powell, A. a Davies, H. (2013) 'What Counts as Good Evidence?.' Llundain: Y Gyngair ar gyfer Tystiolaeth Ddefnyddiol
72. Am fodel o ymagweddau gwahanol ar ansawdd sy'n cynnwys pedwar dimensiwn: (1) ansawdd methodolegol; (2) ansawdd wrth adrodd; (3) priodolrwydd; (4) perthnasedd i bolisi ac arfer, gweler: Boaz, A. ac Ashby, D. (2003) 'Fit for purpose? Assessing research quality for evidence based policy and practice.' ESRC UK Centre for Evidence Based Policy and Practice Working Paper 11.
73. Gweler 'UK Concordat to support Research integrity', a lofnodwyd yn 2012 gan nifer o arianwyr ymchwil allweddol: <http://www.universitiesuk.ac.uk/highereducation/Pages/Theconcordatstosupportresearchintegrity.aspx#.Vd2X5f1VhBc>
74. Callaway, E. (2011) Report finds massive fraud at Dutch universities. 'Nature' 479(15). Gweler: <http://www.nature.com/news/2011/11101/full/479015a.html>
75. Boaz, A. ac Ashby, D. (2003) 'Fit for purpose? Assessing research quality for evidence based policy and practice.' ESRC UK Centre for Evidence Based Policy and Practice Working Paper 11.
76. Trysorlys Ei Mawrhydi, DECC a DEFRA (2012) 'Quality in policy impact evaluation; understanding the effects of policy from other influences.'

77. Trysorlys Ei Mawrhydi (2011) 'The Magenta Book: Guidance for Evaluation.' Llundain: Trysorlys Ei Mawrhydi
78. Trysorlys Ei Mawrhydi (2011) 'The Green Book: Appraisal and Evaluation in Central Government.' Llundain: Trysorlys Ei Mawrhydi.
79. Brooks, G., Hall, J. a Torgerson, C.J. (2006) 'A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling.' Llundain: Adran Addysg a Sgiliau
80. Henderson, M. (2013) 'The Geek Manifesto: Why science matters.' Llundain: Corgi. Tudalen 17.
81. Trysorlys Ei Mawrhydi (2011) 'The Magenta Book: Guidance for Evaluation.' Llundain: Trysorlys Ei Mawrhydi.
82. Gweler Morgan, S.L. (gol.) (2013) 'Handbook of Causal Analysis for Social Research.' Berlin: Springer. Ond mae'n rhaid ei daro ar ei ben: <http://www.tylervigen.com/spurious-correlations>
83. Gweler tudalen 4 yn Stern, E. (2015) 'Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.'
84. Mae ymchwil ansoddol yn archwilio ac yn ceisio deall credoau, profiadau, agweddau ac ymddygiad pobl a sut maent yn rhyngweithio. Mae'n cynhyrchu data nad yw'n rhifiadol. Mae dulliau ymchwil ansoddol cyfarwydd yn cynnwys cyfweiliadau manwl, grwpiau ffocws, dadansoddi dogfennau ac arsylwadau cyfranogwyr, er enghraifft.
85. Er ei fod dros ddegawd oed, mae gan Swyddfa'r Cabinet a hen Swyddfa'r Prif Ymchwilydd Cymdeithasol ganllaw ar ansawdd ymchwil ansoddol. Waeth lle rydych chi'n sefyll yn y rhyfeloedd paradeim, mae digon o gonsensus bod ansawdd yn bwysig, ac na allwn osgoi dod i farn am yr hyn sy'n dda neu beidio. Gweler: <http://www.cebma.org/wp-content/uploads/Spencer-Quality-in-qualitative-evaluation.pdf>
86. Am drosolwg – a beirniadaeth – o hierarchaeth tystiolaeth, gweler: <http://www.alliance4usefulevidence.org/assets/What-Counts-as-Good-Evidence-WEB.pdf>
87. Public Health England (2014) 'Identifying what works for local physical inactivity interventions.'
88. [https://www.nesta.org.uk/sites/default/files/nesta\\_standards\\_of\\_evidence\\_csaif.pdf](https://www.nesta.org.uk/sites/default/files/nesta_standards_of_evidence_csaif.pdf)
89. Dylanwadwyd yn gryf ar safonau Nesta a Project Oracle gan yr Uned Ymchwil Cymdeithasol annibynnol yn Darlington, sy'n gweithio gyda nifer o awdurdodau lleol a darparwyr gwasanaethau plant.
90. Ni Ógáin, E. (2015) 'Impact measurement in impact investing learning from practice.' Llundain: Nesta Impact Investments. Gweler: [http://www.nesta.org.uk/sites/default/files/impact\\_measurement\\_in\\_impact\\_investing.pdf](http://www.nesta.org.uk/sites/default/files/impact_measurement_in_impact_investing.pdf)
91. Ibid
92. Ruiz, F. a Breckon, J. (2014) 'The NICE way; lessons for social policy and practice from the National Institute for Health and Care Excellence.' Llundain: Y Gynghair ar gyfer Tystiolaeth Ddefnyddiol/NICE.
93. Rawlins, M. (2014) Evidence, values and decision-making. 'International Journal of Technology Assessment in Health Care'. 30(2). 233-8.
94. Er enghraifft, gweler tudalen 28 yn Nutley, S. et al. (2013) 'What Counts as Good Evidence?'. Llundain: Y Gynghair ar gyfer Tystiolaeth Ddefnyddiol.
95. I weld y gwahaniaeth rhwng adolygiadau systematig a meta-ddadansoddi, gweler tudalennau 64-5 yn 'The Magenta Book.' Am ganllaw cyflym ar wneud adolygiadau, wedi'i dargedu at lunwyr polisi, gweler dogfen y Gynghair ar gyfer Tystiolaeth Ddefnyddiol (2013) 'Learning from Research: Systematic Reviews for Informing Policy Decisions.': <http://www.alliance4usefulevidence.org/assets/Alliance-FUE-reviews-booklet-3.pdf>
96. Am feirniadaeth ar ddefnyddio cyfodiadau ymchwil mewn polisi, gweler Pawson, R. (2001) 'Evidence Based Policy: In Search of a Method.' ESRC UK Centre for Evidence Based Policy and Practice: Working Paper 3.
97. Gweler tudalen 9 yn <http://www.alliance4usefulevidence.org/assets/Alliance-final-report-08141.pdf>
98. Trysorlys Ei Mawrhydi (2011) 'The Magenta Book: Guidance for Evaluation.' Llundain: Trysorlys Ei Mawrhydi.
99. Thomas, J., Newman, M., Oliver, S. (2013) Rapid evidence assessments of research to inform social policy: taking stock and moving forward. 'Evidence & Policy: A Journal of Research, Debate and Practice.' 9(1). 5-27.
100. <http://www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/dfe-evidence-check-forum/>
101. Begley, C.G. ac Ellis, L.M. (2012) Drug development: Raise standards for preclinical cancer research. 'Nature.' 483. 531-533.
102. <http://www.alliance4usefulevidence.org/replication/>
103. McCormick, J. (2013) 'Evidence Exchange: Learning from social policy from across the UK.' Llundain: Ymddiriedolaeth Carnegie UK a Sefydliad Joseph Rowntree.
104. <http://www.nesta.org.uk/blog/why-what-works-centres-are-working>
105. [https://www.nesta.org.uk/sites/default/files/nesta\\_a4ue\\_and\\_nice\\_paper.pdf](https://www.nesta.org.uk/sites/default/files/nesta_a4ue_and_nice_paper.pdf)
106. <http://www.alliance4usefulevidence.org/a-brief-guide-to-uk-social-policy-databases/>
107. <http://www.scotland.gov.uk/Resource/0038/00385880.pdf>
108. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/243718/evidence-reduce-reoffending.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/243718/evidence-reduce-reoffending.pdf)
109. Er mai canllaw ar gyfer polisi cymdeithasol yn y DU yw hwn, mae'n werth crybwyll canllawiau byr 3ie ar ddatblygu rhyngwladol. Gweler: <http://www.3ieimpact.org/en/evidence/policy-briefs/3ie-briefs/>
110. <http://www.alliance4usefulevidence.org/rapid-evidence-assessments-a-bright-idea-or-a-false-dawn/>
111. <http://www.alliance4usefulevidence.org/the-creation-of-a-new-service-to-unlock-research-expertise-and-you-hold-the-key/>
112. E.e. Y Comisiwn Ewropeaidd (2010) 'Communicating research for evidence-based policymaking – a practical guide for researchers in socio-economic sciences and humanities.' Gweler: <http://www.spia-europa.de/pdf/guide-communicating-research.pdf>
113. Y Cyngor Ymchwil Feddygol (2013) 'Developing and evaluating complex interventions: new guidance.' Gweler: <http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>
114. Beynon, P. et al. (2012) 'What Difference does a Policy Brief Make?.' International Initiative for Impact Evaluation (3ie), Institute of Development Studies, Norwegian Agency for Development Cooperation. Gweler: [http://www.3ieimpact.org/media/filer\\_public/2012/08/22/fullreport\\_what\\_difference\\_does\\_a\\_policy\\_brief\\_make2pdf\\_-\\_adobe\\_acrobat\\_pro.pdf](http://www.3ieimpact.org/media/filer_public/2012/08/22/fullreport_what_difference_does_a_policy_brief_make2pdf_-_adobe_acrobat_pro.pdf)
115. [http://www.3ieimpact.org/media/filer\\_public/2012/08/24/3ie\\_policy\\_brief\\_experiment\\_summary.pdf](http://www.3ieimpact.org/media/filer_public/2012/08/24/3ie_policy_brief_experiment_summary.pdf)
116. Behavioural Insights Team (2014) 'EAST: Four Simple Ways to Apply Behavioural Insights.' Gweler: [http://www.behaviouralinsights.co.uk/wp-content/uploads/2015/07/BIT-Publication-EAST\\_FA\\_WEB.pdf](http://www.behaviouralinsights.co.uk/wp-content/uploads/2015/07/BIT-Publication-EAST_FA_WEB.pdf)
117. Ymchwil wedi'i dyfynnu yn ymchwil y Post Brenhinol (heb ddyddiad) 'Door to Door; Big Return. The seven secrets of highly effective leaflet design.' Gweler: [http://www.royalmail.com/sites/default/files/docs/pdf/seven\\_secrets\\_from\\_rm\\_v1.09.pdf](http://www.royalmail.com/sites/default/files/docs/pdf/seven_secrets_from_rm_v1.09.pdf)
118. Gweler Rousseau, D.M. (ed.) (2012) Oxford Handbook of Evidence Based Management. Rhydychen: Oxford University Press.
119. Allen, G. (2011) 'Early Intervention: The Next Steps An Independent Report to Her Majesty's Government.' Gweler: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/284086/early-intervention-next-steps2.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/284086/early-intervention-next-steps2.pdf)
120. <http://www.theguardian.com/society/2014/may/06/policymakers-neuroscience-justify-early-intervention-agenda-bruce-perry>
121. McCabe, D.P. a Castel, A.D. (2007) Seeing is believing: The effect of brain images on judgments of scientific reasoning. 'Cognition.' 107 (1). 343-352. Gweler: <http://www.imed.jussieu.fr/en/enseignement/Dossier%20articles/article4.pdf>
122. Mae hwn yn rhan o raglen waith Building Capacity to Use Research Evidence (BCURE), wedi'i hariannu gan Adran dros Ddatblygu Rhyngwladol y DU.
123. <http://www.alliance4usefulevidence.org/the-poor-state-of-evaluation-in-government/>





# Introduction

*“There is nothing a government hates more than to be well-informed; for it makes the process of arriving at decisions much more complicated and difficult.”*

John Maynard Keynes

Research evidence can help you understand what works, where, why and for whom. It can also tell you what doesn't work, and you can avoid repeating the failures of others by learning from evaluations of unsuccessful programmes.

Evidence also challenges what we might think is common sense. For instance, it may sound like a good idea to increase the amount of police on the streets to reduce crime or to reduce classroom sizes – but the evidence doesn't necessarily support this. More uniformed police patrolling the streets might make the public feel safer, but it can actually take police away from solving crimes.<sup>1</sup> Despite this, the majority of political leaflets and manifestos in the 2015 UK General Election still claimed that increasing police numbers on the street would reduce crime. Politicians ignored the evidence.<sup>2</sup>

When budgets are tight, we can't afford to waste money on policies and programmes that don't work. Getting laptops and high-tech gizmos into every school may be a good headline-grabbing commitment, but it does little in itself to benefit young people's learning. In fact research by the OECD found that frequent use of computers in schools is often connected with lower academic results.<sup>3</sup>

Throwing money at most problems will of course often do some sort of good. But could that money have been spent more effectively on something cheaper? It seems obvious that reducing classroom sizes would improve teaching and learning quality by boosting the one-to-one attention learners receive. But overall the evidence shows little benefit. To achieve impact, typical classes would need to be reduced to 15 pupils – a costly undertaking for those managing school budgets. Instead there are cheaper and higher impact alternatives, such as giving feedback to teachers or peer tutoring.<sup>4</sup>

Whether it's in a police station, a school classroom or the boardroom of a charity, evidence can help you make better decisions. It is helpful not only in frontline service-delivery, but also in creating smarter organisations – charities, local authorities, government departments – and in developing national policies or charity campaigns.

We have created this guide to point you on the right path to finding what evidence might help you. It should help to build your confidence in 'compiling, assimilating, distilling, interpreting and presenting a strong evidence base'<sup>5</sup> using existing research, and to also think about how you might go on to evaluate your own work.

## Who might read this guide?

This guide is aimed at decision-makers in government, charities, voluntary organisations, professional membership bodies and local authorities working in UK social policy and practice. It is not aimed at trained evaluators and researchers, but instead intends to foster intelligent demand for research evidence from wider audiences.

## How to use this guide

The guide is divided into the following five sections:

### Section A: What is evidence-informed decision-making, and why focus on research?

**Section A** discusses what we mean by evidence-informed decision-making, and why research is an essential element of it.

### Section B: When can evidence help you?

**Section B** explores different scenarios in which using evidence can help you, as well as the types of evidence you might need at different stages of development.

### Section C: 'Horses for courses' – What evidence should you choose?

**Section C** looks at different types of evidence and examines how to choose the most appropriate for your case. It also discusses how to judge the quality of evidence.

### Section D: Where should you look for evidence?

**Section D** offers advice and resources to help you find the right evidence to support your case.

### Section E: How should you communicate your findings?

**Section E** focuses on how to get your message across once you have the evidence you need.

# Section A

## What is evidence-informed decision-making, and why focus on research?

**This section discusses what we mean by evidence-informed decision-making, and why research is an essential element of it.**

To begin, let's be clear about what we don't mean. We are not talking about slavishly following rigid research conclusions. Professional judgement and other sources of information – such as feedback from your stakeholders – will always be important. This guide is not about replacing professional judgement, or pretending that there are easy answers, but about increasing the quality of evidence use in social policy.

A good start in defining what we mean is borrowed from medicine. Two decades ago, David Sackett and his colleagues proposed the following definition that has stood the test of time:

*“Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.”<sup>6</sup>*

This attempt to define Evidence-Based Medicine (EBM) was not the first, but it has been influential and is just as relevant to social policy as it is to medicine. It stresses how research can complement professional judgement or other sources of information.<sup>7</sup>

However, any model of good decision-making should be wary of putting professional judgement on a pedestal as experts can sometimes get it horribly wrong. Later in this section you will read about how we can be 'predictably irrational' and – consciously or unconsciously – make errors in important judgements. In medical practice, one study found that 'cognitive factors' contributed to 74 per cent of mistaken clinical diagnoses – leading to injuries and death.<sup>8</sup> We will explore how to mitigate these errors of judgement in later chapters.

Other decision-making models have also stressed the importance of blending knowledge of evidence with judgement. The UK Civil Service<sup>9</sup> recommends that UK policymakers should have skills on 'politics' and 'delivery' to complement their ability to marshal and apply evidence. This is also relevant to others outside policymaking, such as staff in charities or local authorities. The politics might be slightly different – involving the small 'p' politics of your sector or institution instead of manifestos and political parties – but it is still something to master. Delivery is important for any organisation, but the focus of this guide is on bolstering evidence.



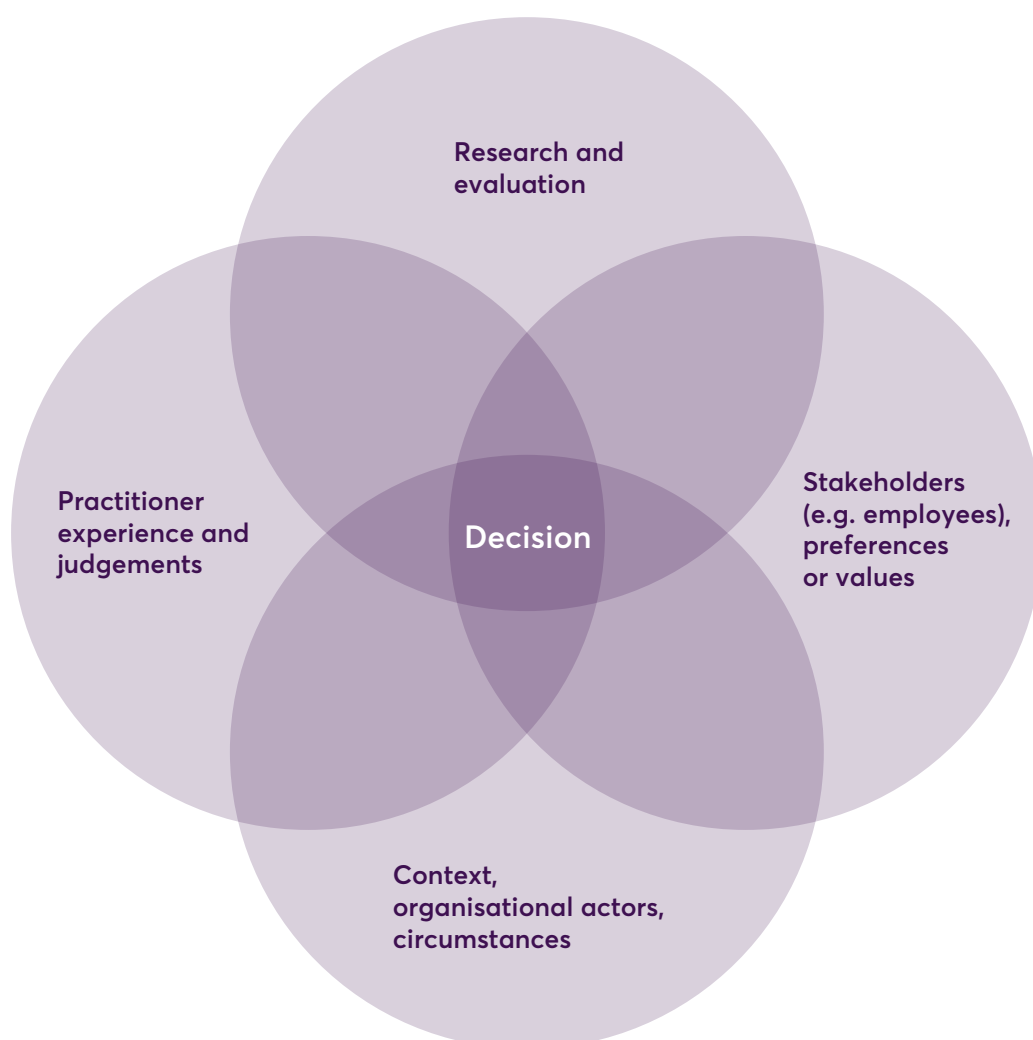
## What is 'evidence' and why do we focus on research?

The Oxford English Dictionary definition of 'evidence' is: *"The available body of facts or information indicating whether a belief or proposition is true or valid."*<sup>10</sup>

We use this dictionary version because many other definitions tend to be rather unhelpful by being overly inclusive – including almost all information – or by being too abstract and vague.

However, we depart from these definitions somewhat by zeroing in on just one type of evidence. Figure A1 shows the different elements that should be part of evidence-based decision-making, but our focus is on the top circle of the Venn diagram: Research and evaluation.

Figure A1: The four elements of evidence-based management



Source: Professor Rob Briner, Centre for Evidence-based Management

The reason for this focus is nicely captured by an earlier guide on evidence use produced during the creation of the ESRC UK Centre for Evidence Based Policy in the early 2000s:

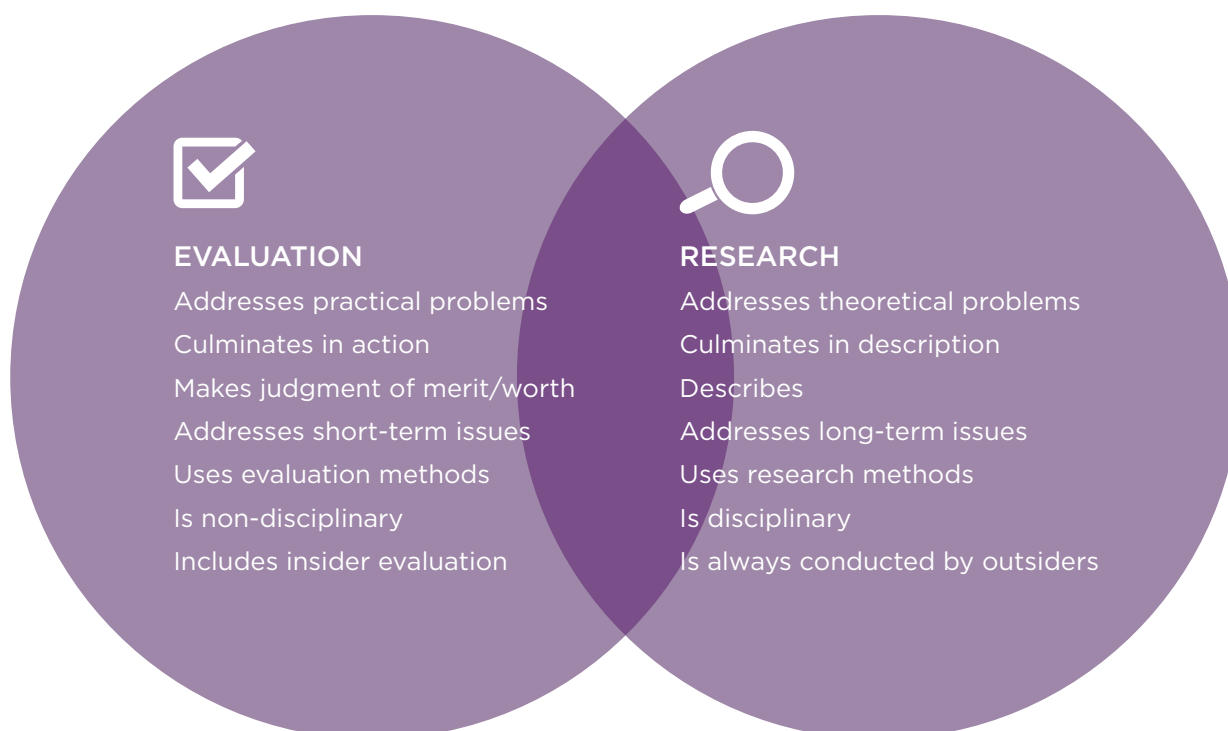
*“When we refer to ‘research evidence’, this includes evidence from published research articles and papers, or unpublished sources such as internally conducted evaluations. Research is only one sort of evidence, but has the advantages of greater **rigour, relevance and independence** when compared to some other types of evidence.”<sup>11</sup>*

So why does research evidence have these advantages? As the authors of the Alliance for Useful Evidence report *What Counts as Good Evidence?* state:

*“The conduct and publication of research involves the explicit documentation of methods, peer review and external scrutiny, resulting in rigour and openness. These features contribute to its systematic nature and help provide a means to judge the trustworthiness of findings. They also offer the potential to assess the validity of one claim compared to another.”<sup>12</sup>*

Other types of evidence – such as in-house evaluations – can be useful and sometimes have these traits. But ‘internal evaluations’ can be hard for others to locate; they are often hidden on organisations’ websites and may not have been peer-reviewed. Too many internal evaluations are also of poor quality (see page 88), and are more likely to be subject to the many biases that can afflict research and evaluation (see page 70).

Figure A2: **Common traits of evaluation versus research**<sup>13</sup>



This guide focuses on research, but there are many overlaps with the field of evaluation and we discuss some approaches to evaluating impact in Section C. We also give most attention to research that deals with impact – whether something has had positive or negative results – as questions on impact are vital to our audience. They are concerned about showing their 'impact' in charities, 'results' in international development or 'what works' for government. The language may change, but the ideal stays the same: to see if you have really made a difference.

We also give prominence to research and evaluation that is ready-made, with no need to run a brand-new study. Many decision-makers are pushed for time and resources and simply can't afford to commission new research. The need for evidence is often now, not in a year's time, so decision-makers require material that can be taken 'off the shelf'. Fortunately, this is possible and we cover this in Section D.

## Experts vs dart-throwing monkeys: the trouble with professional opinion

Another reason we privilege research is because professional judgement can fail to get it right. A study by the American psychologist and political writer Professor Philip E. Tetlock<sup>14</sup> found the predictions of most experts were wrong. He gathered 80,000 expert predictions and compared them to what actually happened. The results were devastating. Academics, government officials, journalists and other pundits performed worse than 'dart-throwing monkeys' in forecasting the future. Indeed, those specialists who had more detailed subject knowledge seemed to perform even worse than average.

## Cognitive bias

But experts are not just bad at predicting the future. They can also make more immediate mistakes. A key insight from psychology is how often people make errors of judgement – not just lay people, but highly-trained surgeons, judges or forensic scientists working in life-or-death scenarios. We carry a 'heavy burden of prejudices, preconceptions and even partiality'.<sup>15</sup> As humans we are 'predictably irrational'<sup>16</sup> and may experience up to 150 cognitive biases that distort our thinking. Cognitive bias refers to our inability to be entirely objective, which may manifest itself via several possible routes – such as perceptual distortion, inaccurate judgments and illogical and/or irrational interpretations. It means that we should be highly cautious about the accuracy of expert decisions.<sup>17</sup>

Even with the best intentions, professionals can get it very wrong. Take for instance one type of cognitive error, confirmation bias. This is the tendency to see the evidence that fits with what we believe, and to ignore or discount what doesn't. Even highly experienced experts can fall into this cognitive trap (see the case study on page 69).

A study of social workers assessing the risk of child abuse 'at the front door' found their accuracy to be 'only slightly better than guessing'. Confirmation bias was one of the factors to blame for this difficulty in decisions on child abuse, according to the Behavioural Insights Team who conducted the study for the UK's Department for Education.<sup>18</sup>



## Forensic scientists fall into the trap of 'confirmation bias'

The 2004 train bombings during Madrid's rush-hour killed 191 people and wounded 1,800. The Spanish authorities suspected al-Qaeda. In the hunt to find the bombers, and under the shadow of 9/11, three FBI fingerprint experts confidently concluded that a print taken from a bag containing detonation devices belonged to Brandon Mayfield, an American lawyer in Oregon.

Mayfield spent 17 days in FBI custody. This was despite constant pleas from Spanish authorities that it couldn't possibly be him. Even after finding out that Mayfield's print was not an identical match to the print left on the bag of detonators, FBI fingerprint examiners still rationalised away the differences. Many pieces of evidence were cherry-picked to support their thinking. Some were not-so-forensic, such as the fact that Mayfield had converted to Islam after marrying his Egyptian wife.

Following this embarrassing episode, the US Department for Justice ordered a full review of the case and ultimately implicated 'confirmation bias' as contributing to Mayfield's misidentification, adding that a *"loss of objectivity"* led examiners to see *"similarities... that were not in fact present"*.

The FBI had repeatedly twisted facts to fit its theory that he was guilty. In fact the FBI's belief that it had its man, despite all contrary evidence, was so strong that it provided misleading sworn statements to a judge.<sup>19</sup>

Being aware of how we can jump to conclusions is important for making us wary of experts. But confirmation bias also highlights how anybody – not just experts – can be highly selective in their use of research evidence. We look for the evidence that fits our beliefs, intentionally or not. This is sometimes referred to as 'policy-based evidence': cherry picking and retro-fitting the evidence to the conclusions we have already fixed in our minds. If we are generous this can be an unconscious mistake, and at worst a deliberate ploy to back up our prejudice by finding the evidence that fits and burying the stuff that doesn't. The Nobel prize-winning psychologist Daniel Kahneman, along with colleagues Dan Lovallo and Olivier Sibony, set out 12 ways to tackle confirmation bias in a useful article in the *Harvard Business Review*.<sup>20</sup>

A similar but different phenomenon is optimism bias. This is a very real danger for anybody setting up a big new project in the public sector or charities – particularly one involving new buildings, civil engineering, IT and outsourcing. One influential study of 258 transport infrastructure projects worth US\$90 billion, found with overwhelming statistical significance that the cost estimates in planning projects are 'highly and systematically misleading'.<sup>21</sup> We often focus too much on the positive, which can lead to major cost overruns.

## Cognitive biases relevant to leaders in social policy and practice



### Optimism bias/Planning fallacy

More likely to falsely expect positive outcomes when planning for the future.



### Confirmation bias

Tendency to interpret and search for information consistent with one's prior beliefs.



### Hindsight bias

Tendency to see past events as being more predictable than they were before the event occurred.



### Loss aversion

Tendency to prefer avoiding losses than to acquire gains.



### Framing effect

Drawing different conclusions from exactly the same information presented in different ways (e.g. would you prefer a ready meal that's '85 per cent fat-free' or '15 per cent fat?').



### The 'availability heuristic'

When people relate the size, frequency or probability of a problem to how easy it is to remember or imagine.



### The 'representativeness heuristic'

When people overestimate the probability of vivid events.



### The 'need for coherence'

The urge to establish patterns and causal relationships when they may not exist.<sup>22</sup>



### Meta-cognitive bias

The belief that we are immune from biases!

(with thanks to Professor Rob Briner, Bath University and Professor Paul Cairney,<sup>23</sup> University of Stirling)

This is not to say that professional judgement is always wrong. Other researchers such as Gary Klein have sung the praises of intuitive expert judgement in his work on 'naturalistic decision-making'. Professional views and gut-instincts can be highly valuable. But we must be aware of their downsides. As Daniel Kahneman asserted in a joint article with Gary Klein in *American Psychologist*, their influential research has flagged that "*professional intuition is sometimes marvellous, and sometimes flawed*".<sup>24</sup>



## The tragedy of professional ignorance over cot death research

Cot death is a horror that can haunt new parents. Fortunately instances of cot death, or to use the medical terminology Sudden Infant Death Syndrome (SIDS), have gone down. But one of the biggest of tragedies of cot death is that if we had looked at the research instead of listening to the experts, many babies' lives might have been saved in Europe, the US and Australasia.

Following the advice of health professionals such as the best-selling Dr Spock, a whole generation of parents laid their babies face-down in the cot believing that they were doing the right thing. But according to Dr Ruth Gilbert of University College London in an article in the *British Medical Journal* in 2008, by 1970 there was significant evidence from clinical research that putting babies to sleep on their front increased the risk of cot death compared with putting babies on their backs.<sup>25</sup> The 'Back to Sleep' public health campaign had a dramatic effect on sudden infant death, but was not launched until November 1991, and the safer sleeping position was not consistently recommended until 1995.

Dr Gilbert believes that the advice to put infants to sleep on their front for nearly half a century was "*contrary to evidence available from 1970 that this was likely to be harmful*". A systematic review (see page 93) of preventable risk factors for SIDS from 1970 would have led to earlier recognition of the harm of sleeping on the front and might have helped prevent some of the 10,000 infant deaths in the UK and 50,000 in Europe, the US and Australasia.

### Key messages from Section A

- We are not advocating slavishly following research at the expense of professional judgement – or other sources of information – but **you also shouldn't put professional expertise on a pedestal**. Research has shown that experts can sometimes get it horribly wrong, and they are not immune from a whole range of social and cognitive biases.
- Creating new evidence can be costly and time-consuming, but there is plenty of **good quality evidence already in existence** that can be taken 'off the shelf'.
- Research is only one sort of evidence, but has the advantages of **greater rigour, relevance and independence** when compared to other types of evidence.

## Section B

# When can evidence help you?

This section explores different scenarios in which using evidence can help you, as well as the types of evidence you might need at different stages of development.

Evidence can make organisations more effective. From more persuasive campaigning to winning grant funding; and from developing a board's decision-making abilities to making sure programmes deliver results – evidence can bolster your work. It doesn't matter if you are a small voluntary organisation or a large government department. Whatever the scale, there's a type of affordable research that can suit your needs.

Below are some examples of when evidence can help you:

Figure B1: Reasons for needing evidence

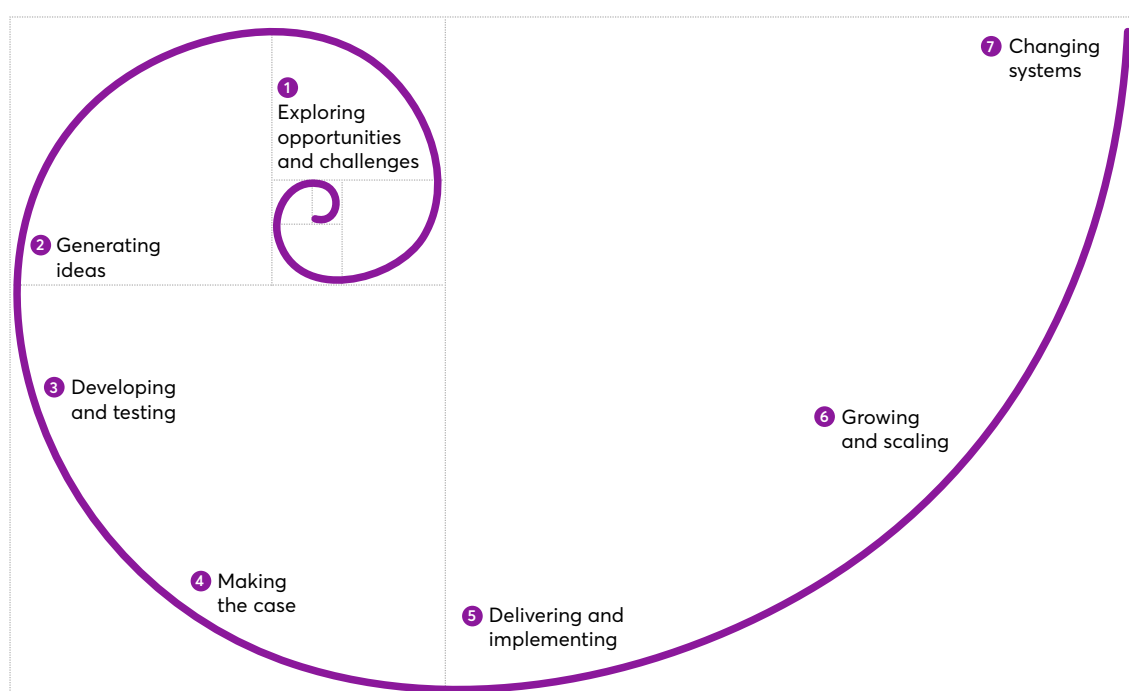




It's a good idea to begin by think about timing. Different evidence will be helpful at different times. You will need to think about the appropriate research to suit different stages of the lifecycle of a new programme, policy or practice.

In the early days of a new initiative, research can identify emerging challenges and the scale of the problem. For instance, a longitudinal study (see Table C:2 on different research approaches) observes people over time – sometimes decades – may highlight the deteriorating public health of a large group of people, such as the depression and morbidity associated with loneliness in old age, or the health and psychological damage of physical inactivity in children.

Figure B2: The Nesta innovation spiral



The spiral above was developed by Nesta to capture the different stages of the innovation process, and can also be used to plot the progress of a new approach to a social issue.<sup>26</sup> Different types of evidence will be needed at the different stages:

### 1. Exploring opportunities and challenges

Working through current research will help you to understand the problems and opportunities around your particular area or issue.

### 2. Generating ideas

After you've identified your focus it's time to hunt for evaluations of interventions and policies that have worked – or failed – in the past. Can you borrow successful ideas from others and avoid interventions that have failed?

### 3. Developing and testing

As new ideas are initiated, it will be time to start thinking about testing and experimenting with different approaches – and about evaluating the impact of these. In Section C, we discuss Standards of Evidence that show the journey new programmes should go on in terms of providing evidence of impact.

#### 4. Making the case

Having evidence of the impact of your testing will then help with making the case to funders and supporters. This will put you in a stronger position to persuade your backers and move on to delivery and implementation.

#### 5. Delivery and implementation

Once you have implemented a policy, programme or project you need to think about how you can evidence your impact. This will help take you from a position of rhetoric and saying 'trust me, this project is working' to one based on more trustworthy evidence. It will also help with your accountability – by showing that you are making a difference, that you are value for money, and have opened up your work to evaluation.

#### 6. Growing, scaling and spreading

The holy grail for many innovations is to replicate and grow so that they change things on a bigger scale – and can potentially be successfully copied in other locations (see Nesta's report *Making It Big: Strategies for scaling social innovations*<sup>27</sup>).

#### 7. Changing systems

Achieving system-wide changes is extremely complex, but can be seen in the mainstreaming of recycling, the hospice movement or the ban on smoking in public places. A good way to reach this scale is to have multiple independent studies to show that something is not just a flash in the pan but an effective approach that works in many places.

### Adopting ideas from other regions or sectors

Even if you find evidence of success of a policy or practice, would adoption of those ideas work in your area?

For instance, there seemed to be plenty of positive research to support the Nurse Family Partnership (NFP), a US-created programme that uses specially trained nurses to carry out home visits with teenage first-time mothers from low-income families. It has shown some astonishing success, with over two decades worth of evidence to back up its effectiveness. But much of that research comes from the US, so would it work in other countries?

It certainly doesn't seem to be the case in the UK. An evaluation found that NFP didn't seem to translate and wasn't as effective in the UK – perhaps because through the NHS, British mothers already had more support than US ones.<sup>28</sup> This example shows that context matters, and that we should not cut-and-paste programmes from elsewhere, unless they are adapted to local circumstances.

And when designing new programmes aimed at changing how we act – for example, cutting smoking or reducing anti-social behaviour – it's always worth checking out research from those in the fields of social psychology and behavioural research, such as The Behavioural Insights Team. There is a range of techniques to 'nudge' people towards desired actions based on robust research such as randomised controlled trials or ethnographic research (see Section C on research methods and approaches).

## Creating a Theory of Change

In the early stages of any intervention, it's important to logically describe what you do and why it matters coherently, clearly and convincingly. This is often referred to as a Theory of Change, and aims to give a:

*"...clear, concise and convincing explanation of what you do, what impact you aim to have, and how you believe you will have it. It is a vital foundation of any programme, and a prerequisite for effective evaluation."<sup>29</sup>*

A Theory of Change is a useful way to be more explicit about what evidence you are using from others – and to be clearer about how you are going to get results. For instance, if you want to run a programme to get kids more physically active, what evidence is out there on effective behavioural change? Is building more playgrounds enough? What incentives really work on children (and their parents) to get them off the couch? You could get some useful pointers from behavioural science or the evaluation of other physical activity programmes.

A Theory of Change helps you be explicit about your goals – and how you'll achieve those goals. It helps to avoid just 'winging it', hoping that your new innovation may stick and that your assumptions are correct. Another benefit of doing a Theory of Change is that it's a first step in designing an effective evaluation as it tries to accurately identify all of your outcomes that then need to be measured.

For more on policy development, rather than growing charities or programmes outside government, see Appendix 1 for rationales for evidence and types of evidence required to match your policy questions.

The most important message here is to think about appropriateness. You want to find the research that fits your needs and the stage of development that you are at. The sort of evidence you'll require in the first few months of a new programme will be very different from what you need when you are more established. We revisit this crucial issue of appropriateness in the next section.

## Key messages from Section B

- There is a **wide range of situations in which evidence can help you**. It's not just the obvious ones around capturing the results of programmes, policies and practice. There may be other benefits that you haven't thought about, such as creating more persuasive campaigning, winning grant funding or stopping doing something that's not working.
- You need to think about the **timing and appropriateness of evidence**. Think where you may be on the innovation spiral – and what evidence may suit your needs. For instance, in the early days of developing a new innovation you'll most likely want to avoid doing a large, costly evaluation of your own work and instead learn from others through past evaluations or wider social and scientific research.
- Early stages of innovation are also the time to create a **Theory of Change**. This is a useful way to be clearer about what evidence you are using and how you are going to get results. It will also help you to design an effective evaluation.
- Programmes, policies and practices that are **more established can set up their own experiments or evaluations of impact**. As an innovation grows, you should also consider multiple replications to check that success was not just a flash in the pan, but that it can work in other places and contexts.
- **It doesn't matter what scale of organisation** you are. There is now a lot of freely available research even if you are a small voluntary organisation.

## Section C

# 'Horses for courses' – What evidence should you choose?

**This section looks at different types of evidence and examines how to choose the most appropriate for your case. It also discusses how to judge the quality of evidence.**

Not all evidence is equal. Some is stronger – and more relevant to your challenge – than others.

A key message of this practice guide is to think about appropriateness. Too many people get into trouble by not thinking clearly about what sort of research design, method or way of collecting data is really going to meet their needs. We can be 'methods-led'.<sup>30</sup> In other words, we pick our pet approach, such as questionnaires or RCTs, because that's what we're most comfortable with rather than being more open to the breadth of methods – and thinking which ones are best suited to answer our challenge.

For instance, experimental research is more suited to evaluating impact and 'what works'. The UK network of 'Sure Start' children's centres – which provide help and advice on child and family issues – were informed by randomised control trial evaluations and systematic reviews (see page 82 for more on these approaches) of early years initiatives in the US.<sup>31</sup> But other types of research can help give other insights. For instance, policymakers setting up Sure Start were inspired by longitudinal studies – research tracking people over decades – that showed the long-term trends of children's poverty.

Other sorts of research may not look so much at impact, but can reveal why and how things are working.<sup>32</sup> For instance, Sunderland City Council found that using ethnography<sup>33</sup> – where the researcher observes from the point of view of the subject – helped Sunderland redesign its employment and welfare programme. The council was able to get closer to its customers' needs by putting itself in the shoes of citizens through ethnography. They were then able to make major savings by getting citizens back into sustained work (see case study on the following page).



## Learning by doing – ethnography and what works in finding a job

Back in 2007, more than a quarter of the working age population in Sunderland were economically inactive and Sunderland City Council was spending a significant amount on benefit claims. But very little progress was being made, so the council tried a new approach; they carried out ethnographic research to unpick the real story of people's often difficult and bureaucratic journey back to work. Ethnography allowed them think about services from a user perspective: through travelling with and talking to people on that journey, they were able to design a suite of services that supported them through it.

In its initial phase the Make It Work programme supported more than 1,000 people, generating early savings of more than a quarter of a million pounds for the council. As well as using ethnography, their approach was experimental, informed by rigorous evidence and tested by iterative prototyping – checking if things were working, and learning quickly and cheaply from mistakes and successes.

What this example stresses is that we don't always know what works – there are gaps and uncertainties in the evidence. But we can take an experimental approach and test as we go, rather than unrealistically thinking there will be a fully-formed solution ready for implementation.

The programme's success depended on Sunderland Council's willingness to try something new using evidence based on local user needs, to fund it properly and to give it the space and permission to experiment. The total cost of running the programme was £180,000. An economist for the council has estimated overall cost avoidance for the council of £435,000 through participants entering sustained work, amounting to an initial saving of approximately £255,000.

This early return on investment is dwarfed by the long-term savings of reducing worklessness. The current Minister of State for Welfare Reform at the Department for Work and Pensions has estimated that it is economically rational to spend £62,000 on supporting the average unemployed person back into work. The average cost of participation in Make it Work is only £5,000.

*Source: Radical Efficiency; different, better, lower cost public services, Nesta 2010*

Whatever the method, whether it is trials or ethnography, the type of research needs to fit the needs of the challenge – in other words: 'horses for courses'.<sup>34</sup>

## PART 1: Different types of research methods, designs and approaches

So how do you go about understanding which research method is appropriate for your case? If you are not a researcher, it can be daunting to come face-to-face with the jargon and endless lists of approaches.<sup>35</sup> See for instance these examples:



Adapted from Luff *et al.* (2015)

But such long lists shouldn't put you off as you don't need to know every type of approach. For the non-specialist 'consumer' of research, it's more important – and easier – to understand the assumptions that underpin these ways of doing research.<sup>36</sup>



These assumptions are captured by a Department for International Development (DFID) *How to Note*<sup>37</sup> on assessing evidence – and are just as relevant to social policy as international development:

- Some research designs are better suited for **demonstrating** the presence of a causal relationship, such as experimental and quasi-experimental designs.
- Others are more appropriate for **explaining** such causal relationships – see *Beyond experiments: new approaches to evaluation* on page 85.
- While some designs are more useful for **describing** political, social and environmental contexts – such as observational studies.

To understand a bit more about the pros and cons of these approaches, see Table C.1.

**Table C1: Different designs, methods and approaches to research evidence – a brief overview**

Types of research and evaluation	What is it?	Pros	Cons
<b>Personal anecdote</b>	An account by a person with direct experience of, or affected by, a particular issue.	Powerful and immediate; may give vivid insights into events concealed from much of the population. <sup>38</sup>	Difficult to verify; may lead to inflation of prevalence; emotive first-person stories may inhibit critical appraisal; individual anecdotes may not be representative.
<b>Service use feedback</b>	Narrative or survey accounts of user views or reported needs.	Valuable insights from those at the receiving end; compels professionals to stay focused on the service users' priorities.	Correlation between satisfaction and service effectiveness is low; expressed needs may not translate into actual service use.
<b>Single case study</b>	Detailed and intensive examination of a single case such as community, family, organisation, event or individual.	Easy and practical; can be used by practitioners and non-specialists; good for interventions that have already happened; can possibly identify adverse consequences from intervention; helps to describe innovations; generates hypotheses for other studies.	Bad at inferring causations; small size means hard to generalise to national/ population level.
<b>Case control studies</b>	Compares a group who have, say, a health condition with a group of people that do not have it, and looks back in time to see how the characteristics of the two groups differ.	Require fewer resources to carry out than intervention studies; useful when RCTs not practical, e.g. studies of cot death.	Rare in social policy (see closely related 'case-based' evaluation below for approach more common in social policy), more frequent in epidemiology and health; provide less evidence for causal inference than an RCT; high risk of bias e.g. recall bias, selection bias, interviewer bias.

<b>Participatory</b>	An approach where the judgements and experiences of stakeholders and beneficiaries are highlighted. May be described as normative designs; participatory or democratic evaluation; empowerment evaluation; learning by doing; policy dialogue; collaborative action research.	Beneficiaries are best able to identify the most relevant theories of change and meaningful outcomes; more potential to be ethical and democratic and understand what beneficiaries really need; more appropriate to 'values-led' interventions; more opportunities for programme adoption due to closer ties to beneficiaries; ability to adapt and customise intervention.	Argument that it is fundamentally un-objective, open to bias and not really research but more about ideology and community activism. <sup>39</sup>
<b>Theory-based</b>	An approach to evaluation that looks at what happens compared with pre-existing theories or causal pathways identified during an evaluation. Can be associated with realist evaluation; qualitative comparative analysis (QCA); contribution analysis; process tracing.	Strong on explanation of causes; can be used in messier areas of social policy where there may be many causes and context is important. <i>[Note: this category is used very broadly here, for more detail on this area see Stern (2015)<sup>40</sup> and White and Phillips (2012)<sup>41</sup>].</i>	Lack of agreed methods; opportunities for bias; weak on estimating quantities or extent of impact; relatively little evaluation of the techniques used e.g. compared to the large body of literature criticising RCTs.
<b>Cross-sectional study</b>	A representative sample of people surveyed at one point in time. Although surveys such as questionnaires and structured interviews are commonly used in cross-sectional design, they are not the only way e.g. content analysis or analysis of official statistics can be used.	Quantitative data can be examined to detect patterns of association; relatively cheap and ethical; survey can be repeated at intervals, illustrating changing trends over time (see Longitudinal below).	Establishes association at best, not causality; rarely possible to attribute any measured change to the intervention, or to understand what would have happened in the absence of the intervention – e.g. change could have been to broader issues such as economic conditions, weather, media campaigns – and not the intervention. Other disadvantages: risk of recall bias, social desirability bias, researcher's (Neyman) bias; group sizes may be unequal; confounders may be unequally distributed.
<b>Cohort/ Longitudinal studies</b>	The same sample of people surveyed over several points in time, sometimes from childhood to old age.	Best source of evidence on association between childhood experience and adult outcomes can give powerful support for certain early interventions.	Data often emerges too late for effective policymaking; study members may drop out over time; very expensive approach when maintained over decades.

<p><b>Quasi-experimental design</b></p>	<p>Different interventions are offered but with no random allocation to groups, i.e. through the use of natural populations or case matching.</p>	<p>Can provide reasonably strong evidence of the relationship between the intervention and the measured outcomes; powerful method of exploring the impact of an intervention when randomisation is impossible; can be applied to large communities as well as groups; no need for randomisation from the start (ex-ante), which avoids the PR and ethical issues of randomisation.</p>	<p>Inability to ensure equivalence of groups and to prevent change over time can result in less reliable findings; matching techniques tend to require a lot of data in both intervention and comparison groups which can be time-consuming and expensive; a good understanding is required of the factors that need to be matched – without this, it remains possible that there are systematic differences between the two groups that are not being controlled for; these designs require complex analytical work and specialist knowledge.</p>
<p><b>Randomised control trial (RCT)</b></p>	<p>One group receives an intervention while another receives none or one of another type, with the chance of trial (RCT) being allocated to either group being identical.</p>	<p>Offers the most robust, reliable findings, which give confidence that any measured difference between groups is the result of the intervention; random allocation should overcome any systematic difference between groups; greater confidence in the <b>effect size</b>, and the relationship between the intervention and outcome; internationally recognised approach.</p>	<p>Poor on taking context into account e.g. cultural, institutional, historical and economic settings; difficult to conduct at a national population level; risk that when used at small pilot level not relevant to national/population level (although this is a risk for all designs); can be hard to manipulate variables to experiment in social policy e.g. class, race or where you live; mistakes in randomisation can invalidate results; can be hard to persuade decision-makers (e.g. politicians) of benefits of this design; potential political, ethical and PR issues over randomisation (e.g. some groups randomly getting potential beneficial intervention, and not others); can take more management time and longer to set up than quasi-experiments.</p>
<p><b>Systematic reviews/ meta-analysis</b></p>	<p>Aggregation of results from eligible studies, with the eligibility criteria defined in advance and methodologies reviewed.</p>	<p>Best source of reassurance that an intervention works (or doesn't); meta-analysis pools statistical results; large reviews carry considerable statistical power; is replicable by other researchers; can be applied to any kind of data.</p>	<p>Requires a substantial number of robust primary studies in a given area; methodology less well developed for synthesising qualitative data and 'grey' literature [<i>For wider approaches to synthesis, see other approaches such as narrative synthesis, and realist-based synthesis</i>].</p>

Adapted from: 'Quality in policy impact evaluation: understanding the effects of policy from other influences.' (supplementary Magenta Book guidance) HM Treasury/DEFRA/DECC (HM Government: 2012); 'The Evidence Guide: Using Research and Evaluation in Social Care and Allied Professions.' (2006) Barnado's/What Works for Children?/Centre for Evidence-Based Social Services; and Petticrew, M. and Roberts, H. (2003). Evidence, hierarchies and typologies: horses for courses. *Journal of Epidemiology and Community Health.* 57: 527-529 (2003); Stern, E. (2015) 'Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.' Table 2, p18.

## Experimental research – why all the fuss?

Experimental research such as randomised controlled trials (RCTs) have received a lot of attention<sup>42</sup> – some might say too much. So why have they attracted so much interest?

Frequently those working in social policy want to know if an intervention has had an impact. Has there been a causal link between your new programme and the 'outcome', 'impact', 'effect' and 'result' at the end? While they are often hard to implement in practice, experimental designs have a somewhat better chance of tracking this cause and effect.

Experiments can simply test two non-randomised groups before and after an intervention. But there may be a risk that the groups are biased or not representative of the groups under investigation – for example, somebody who asks for intervention may be more predisposed to positive outcomes. Perhaps they are more motivated, healthier, confident, and thus more likely to do well regardless of any 'treatment'.

To get rid of this bias, we can use randomisation: the groups compared with each other are selected entirely randomly, for example by drawing lots. The box below gives a fuller description of this approach, taken from a supplement to HM Treasury's *Magenta Book*.

### Experimental research designs

"An experimental design, conducted properly, will establish whether an intervention caused an outcome. Such evaluation designs use random allocation to assign units of assessment (individuals/groups) to either the intervention or counterfactual group (often called 'control group' in experimental design). Given appropriate sample sizes and appropriate allocation to experimental or control groups, this is the strongest form of design for an impact evaluation, as the random allocation minimises the likelihood of any systematic differences – either known or unknown – between the groups. It therefore allows for an attribution of cause and effect."

Taken from *Magenta Guide: HM Treasury, DECC and DEFRA (2012) 'Quality in policy impact evaluation; understanding the effects of policy from other influences.'*

The introduction of a control group eliminates a whole host of biases that normally complicate the evaluation process – for example, if you introduce a new 'back to work' scheme, how will you know whether those receiving the extra support might not have found a job anyway?



### Making it work – experimental research on incapacity benefit

In 2003, the Department for Work and Pensions (DWP) conducted a randomised controlled trial to examine the impact of three new programmes for Incapacity Benefit claimants. These were: support at work, support focused on their individual health needs, or both. The extra support cost £1,400 on average, but the RCT found no benefit over the standard support that was already available. It ultimately saved the taxpayer a lot of money as it provided unambiguous evidence that the costly additional support was not having the intended effect.

More recently the DWP was keen to explore whether the intensity of the signing-on process required of jobseekers on benefits could be reduced without worsening outcomes.

In a trial involving over 60,000 people, the usual fortnightly signing-on process was compared against several others that were less resource intensive (e.g. signing-on by telephone, less frequently). All of the alternatives to the status quo that were tested in trials large enough to show reliable effects were found to increase the time people took to find work. As a result, despite other changes to the benefits system, DWP policy continues to require people to sign on on a fortnightly basis.

Source: Cabinet Office (2012) *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*.

## Practical and ethical issues

A common criticism of RCTs is that they are impractical, but there are often ways round this. For example, the so-called 'stepped wedge' design (see Appendix 2 on different ways of doing trials) can be used as a solution to some of the ethical and practical problems. In this approach, everybody does, eventually, get the intervention – such as new learning approach in schools or a crime-reduction innovation.

But, with stepped wedge trials, the order in which participants receive the intervention is determined by lottery. It is gradually phased in over time – but who gets it is chosen at random. This can be highly practical in a world of austerity. As there is frequently not enough resources to do a full roll-out of a new programme all in one go, a gradual roll-out can be cost-effective. Appendix 2 has a short overview of this and four other different experimental research methods.

## Useful guides on experimental research

- *Randomised controlled trials – gold standard or fool's gold? The role of experimental methods in voluntary sector impact assessment* – a ten-page basic introduction to RCTs for charities by NCVO/Charities Evaluation Service.
- *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials* – a helpful guide from the Cabinet Office.
- *Better Public Services through Experimental Government* – a Nesta report on some of the myths around the ethics and costs of RCTs.
- *Developing and evaluating complex interventions: new guidance* – excellent guidance on a variety of RCTs and other ways of evaluating complex interventions from the Medical Research Council (with examples from social policy, not just medicine).

## When randomisation is not appropriate

Randomly choosing your subjects for research may not be appropriate. It may be completely unrealistic to deliberately expose your experimental 'subjects' to something harmful such as cigarette smoking or a crime-ridden housing estate. It may be more appropriate to use non-randomised designs or 'natural experiments'.

The cot-death case study (see page 71) is an example where nobody in their right mind would run an experiment – with some babies sleeping on their fronts and others on their backs to test if they die or not. But we can still compare groups that have had different experiences. The research on cot-deaths used 'case controlled studies'<sup>43</sup> (see item six in Table C1).

There are other examples of research that have also bypassed traditional RCTs but still used comparable groups to see if there is some link. For example we have been able to compare different cases on legislation that restrict access to the means of suicide, the impact on air pollution controls in different areas and the impact of banning smoking in public places.<sup>44</sup>

## Quasi-experimental designs

There are a wide variety of 'quasi-experimental designs' (QEDs), often with cumbersome titles such as Interrupted Time Series Designs, Proxy Pre-test Design, Non-equivalent Dependent Variables Design, Pattern Matching Design and the Regression Point Displacement Design.

The influential proponent of RCTs and QEDs in the 1960s in the US, Professor Donald Campbell, often referred to them as "queasy" experiments because they give experimental purists a queasy feeling. But QEDs are still helpful when RCTs are not feasible. The Alliance for Useful Evidence has worked with the European Commission (EC), JPAL Europe and the LSE to produce *Testing Social Policy Innovations*<sup>45</sup> that covers many common quasi-experimental designs, available on our website.

## Beyond experiments: new approaches to evaluation

A lot of current thinking on impact has moved away from sole dependence on experiments.<sup>46</sup> RCTs or quasi-experiments may work well when there is a simple intervention that can be tested. However, rarely do we have such simple interventions. NGOs working in civil society, for instance, rarely work alone or have the chance to manipulate a clear experimental 'treatment' for an RCT. Evaluators are looking at other ways of doing things, approaches that involve moving beyond talking of simple one-to-one causality – and thinking about your 'contribution'.

In the complex world of social policy, it's unlikely that your programme is the necessary or sufficient condition for success. It's likely to be just one factor among many, part of a 'causal package'. Programme success depends on what else is going on or has gone on around you. In the past, a simple causal explanation of 'what works' may have sufficed but nowadays it is now much more common for evaluation researchers to ask: 'did we make a difference?'<sup>47</sup>

The case for such theory-based approaches is convincingly set out in an excellent guide produced for the Big Lottery Fund, BOND, Comic Relief and DFID. It shows how explanation, not just causation, is important for any evaluator:

*"You might draw a conclusion (or causal inference) from an evaluation that funding for education programmes for girls led to or 'caused' higher family income in a particular community. However, when it becomes evident that similar educational programmes do not always lead to the same result in all places, people start to ask 'why?'.*

*In IE [impact evaluation], as in scientific research, explanation ultimately relies on good theories. Opening up the 'black box' that connects 'causes' and 'effects' requires different kinds of analysis, which is what 'theories of change' and 'programme theory'... are intended to support.*

*Developments in IE have also made evaluators aware that they need to draw on broader community, social and economic theories in order to interpret complex and often confusing or even contradictory data."*

Source: Stern, E. (2015) *Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector*. Prepared for Bond, Comic Relief, Big Lottery

There are a range of approaches that are becoming more popular with evaluators, such as theory-based<sup>48</sup> and case-based designs, realist evaluation,<sup>49</sup> contribution analysis,<sup>50</sup> process tracing<sup>51</sup> or qualitative comparative analysis.<sup>52</sup>

Rather than overgeneralising about these methods, it's worth touching on one example: contribution analysis. This form of research does not attempt to prove that one factor – say a policy – 'caused' the desired outcome, but rather to explore the contribution a policy is making to observed results. It recognises that effects are produced by several causes at the same time, not just one. The Scottish Government has a short guide<sup>53</sup> on this technique and says that it can be useful when experiments are not practical:

*"By developing a 'theory of change' showing the links between the activities, outcomes and contexts of the policy and collecting evidence from various sources to test this theory, the aim is to build a credible 'performance story'. This can demonstrate whether policies and programmes were indeed an important influencing factor in driving change, perhaps along with other factors."*

These approaches may not all be new.<sup>54</sup> Indeed, having a good theory has arguably always been at the heart of good science – so 'theory-based' approaches don't sound too novel. But these ways of evaluating impact have grown in popularity and they do help evaluators address multiple causality.

## Was it worth it? The value of cost-benefit analysis

If you throw lots of money at any social problem, you are likely to get some benefit. But is it value for money? When budgets are constrained, we need to make hard financial decisions about whether one thing is more valuable compared to another. Policymakers, commissioners of services, grant-making bodies and charitable funders alike are asking for more and better impact measures for every penny they disburse. Techniques such as cost benefit analysis (CBA) can help create such insights on value. It results in a financial value being ascribed to impact, which means that in the current funding climate many charities and public bodies are understandably intrigued.<sup>55</sup>

It's important to note that there is a range of techniques. A cost benefit analysis (CBA) is only one way to compare costs with financial benefits. Two other approaches are worth mentioning: cost effectiveness analysis (CEA) and cost consequence analysis (CCA). The differences between CEAs and CBAs are quite technical and are one for specialists, but the general difference with CBAs is that they measure results in units rather than monetary figures. Other techniques have also been developed specifically for social and environmental value such as social return on investment (SROI).<sup>56</sup>

Because of this broad range of approaches, if you want to run a CBA on one of your own programmes of work we recommend that you use a tool that is already widely used and accepted – see below for a selection of guides on cost benefit analysis.



## Guides on cost benefit analysis

The HM Treasury's *Green Book*<sup>57</sup> – a guide for the appraisal and evaluation of policies, programmes and projects that is relevant for charities and service providers as well as government – features a CBA model frequently used in the public sector.

Organisations wishing to develop this sort of analysis need to collect data on elements such as key performance indicators, costs of service delivery, and (where possible) outcomes or impact on service users. Guidance on what data to collect can be found in another UK government publication entitled *Choosing the Right FABRIC*,<sup>58</sup> and many organisations operating in the public sector have taken an approach to the collection and analysis of performance data informed by this document.

Another useful guide, particularly for those working on local projects and who have tight analytical budgets, is *Supporting public service transformation: cost benefit analysis guidance for local partnerships* produced by HM Treasury, Public Service Transformation Network and New Economy in Manchester.<sup>59</sup> It is designed to simplify and lower the cost of performing CBAs for local programmes aimed at improving public services where analytical and research resources may be relatively limited. It's relevant to charities and business, as well as public sector.

For charities, the Big Lottery Fund's *A Guide to Cost Benefit Analysis' Wellbeing Programme: Evaluation and Learning* will also be useful.

## Key messages for Part 1 of Section C

- **Not all evidence is equal.** Some is better quality and will be more appropriate to your challenge.
- Whatever the method, whether it is RCTs or ethnography, the type of research needs to **fit the needs of the challenge** – in other words: 'horses for courses'.<sup>60</sup>
- Don't **be put off by long lists of research methods**, designs and approaches. It's more important to understand the assumptions that underpin these ways of doing research.
- Some research designs are better suited for demonstrating the presence of a **causal relationship**, such as experimental and quasi-experimental designs.
- Other research approaches are more appropriate for explaining such causal relationships. **Theory-based evaluation** and techniques such as contribution analysis are increasingly popular with evaluators. These techniques can be helpful when it's hard to experiment or impossible to attribute your single policy or programmes to any single clear result.
- It's not just about whether your intervention worked or not, but whether it was value for money. Use techniques such as **cost benefit analysis** to understand the financial value of your impact.

## PART 2: How do you judge the quality of research?

Another way to help you choose which sort of research you need is to ask a different question: what research can you trust? What is good enough evidence to fit your needs? Perhaps it is only the academic top-ranking journals like *Nature*, *BMJ* or the *Quarterly Journal of Economics*. But what about the many in-house evaluations conducted by charities, local authorities and government? Surely they too must have a place on the evidence-table?

Being in a peer-reviewed research journal is one way to help you feel confident of the research, but it's no guarantee. In a famous paper, John Ioannidis from Stanford University caused a stir by arguing that 'most published findings are probably false'. He examined the most cited papers (1,000+ citations) in the best regarded medical journals in the world – largely drawn from *The Lancet*, the *New England Journal of Medicine* and the *Journal of the American Medical Association*.<sup>61</sup> Of those with claims of efficacy whose results had been tested, 41 per cent were either found to be wrong, or the impact was much smaller than the original study suggested. The Ioannidis study is a decade old, but in 2013 he told the International Congress on Peer Review and Biomedical Publication that the problem had not gone away.<sup>62</sup>

### Peer-reviewed research: the decision-maker's comfort zone?

We must also be conscious that peer review – the cornerstone of academic journals – is far from perfect. There can be unconscious biases such as 'herding', where the behaviour of reviewers is influenced by the behaviour of their peers. And the whole setup can be skewed by 'publication bias': positive results have a better chance of being published,<sup>63</sup> while negative data gets shelved. A survey by Stanford University found that most 'null studies' in sociology never see the light of day: just 20 per cent had appeared in a journal, and 65 per cent had not even been written up.<sup>64</sup>

This could have serious consequences for decision-makers reviewing evidence. If you never see all the negative studies on a new intervention, you may wrongly conclude that all is well. In the clinical health field, hiding the results of clinical trials of drugs like Tamiflu may literally kill.<sup>65</sup>

Nevertheless, despite these problems, peer review is still, for most, the 'gold standard': a check on bad work.<sup>66</sup> Using a journal that has been peer-reviewed by other experts is one way of helping you be somewhat more confident that you can trust the research. It still can take years before research gets published and that can be too long if you have to make a quick decision, but peer-review has some quality controls to help you feel confident.

If the research hasn't gone through a peer-reviewed journal, there may be other channels. For those working in government, it's recommended that you consult learned societies (such as the Royal Society, Royal Statistical Society, British Academy and others) or scientific and expert committees as sensible buttresses of quality.<sup>67</sup> In the charity sector, bodies like **Evaluation Support Scotland** have advised using peer networks and talking to other like-minded organisations.<sup>68</sup>

## Nailing down a definition of 'high quality' research

Peer review may give us some modicum of comfort. But what do you do if you are going to include wider evidence that hasn't been checked by other experts? How do you decide if it is up to scratch?

It would clearly be daft to miss important 'grey literature' by research bodies such as the Royal Society for Edinburgh, King's Fund or the Institute for Fiscal Studies just because it wasn't in a peer-reviewed journal. Or to miss the rich seam of government or charity evaluations that never see the 'light' by being published in academic outlets. But it feels a daunting task after reading the National Audit Office's damning report on evaluation.<sup>69</sup> Looking at central government, it found that less than half of the evaluations they examined had provided robust evidence on policy impact – meaning that the findings could not be relied on.

A good start in trying to vet quality is defining it. One of the problems, however, is that phrases such as 'quality', 'standards', 'robustness' and 'strength' are bandied around as if they were interchangeable, and without clearly defining what they are. This makes for a lot of messiness and misunderstanding.

For instance, in some guidance,<sup>70</sup> research 'quality' means using particular types of design and method – such as RCTs. This focus on quality is a deliberate nod to some of the formal clinical and health approaches to assessing evidence quality, such as the GRADE or Scientific Maryland Scale system.<sup>71</sup>

But for others, quality can mean good practice in how you report<sup>72</sup> your research, or your integrity<sup>73</sup> in not faking your data – as seen in the academic scandal of Dutch psychologist Diederik Stapel, who fabricated data for years and published it in at least 30 peer-reviewed papers.<sup>74</sup> Even more confusingly, some researchers have judged quality as being about utility – how relevant your work was to policy and practice.<sup>75</sup> While utility is a commendable goal, it may sound confusing to non-specialists to equate quality with usefulness.

However, for most of the literature, discussions of research quality have zeroed-in on methodological designs. Particular methods like RCTs have won this privileged position as 'high quality'. According to a supplement to the HM Treasury's *Magenta Book*<sup>76</sup> – an evaluation 'bible' – experimental and quasi-experimental designs get closer to seeing if interventions 'cause' outcomes (see a short description of these designs in Appendix 2).

### HM Treasury's *Magenta Book* and *Green Book*

When thinking about what is strong evidence, it is impossible to ignore two key HM Treasury tomes of guidance – the *Magenta Book*<sup>77</sup> and *Green Book*.<sup>78</sup> While both are hugely long – 141 pages and 118 pages respectively – they are readable (all-things-being-relative) and avoid much technical language of economics and evaluation. They are essential reading for anybody making judgements on the strength of evidence.

While both have value for vetting evidence, the *Magenta Book* is particularly helpful in planning your own evaluations or reviewing evidence created by others, whereas the *Green Book* is more future-orientated and perhaps more for officials in central government – offering help on, for instance, how to build a good business case for a new policy.

The *Magenta Book* was also written with wider audiences in mind than Whitehall and so is highly relevant to local authorities, charities, public service providers or external consultants and evaluators that want to work with government.



## The danger of using 'weak' experiments to inform policy – the case of synthetic phonics

The current policy for teaching children to read is based on synthetic phonics, by which children learn to match sounds to letters and groups of letters. A team commissioned by the Department for Education and Skills and led by Professor Carole Torgerson, then at York University, was asked to review the evidence on synthetic phonics.

Although there was some promising evidence from the US and a small-scale study in Clackmannanshire, Scotland, showing that the approach worked, the evidence was limited. Their government review found only a dozen small experimental trials, the biggest of which involved 120 children.<sup>79</sup> They urged caution in making national policy. Torgerson recommended in her report that the government should roll it out gradually, with the first areas to benefit to be chosen at random in a trial.

But this advice was ignored, according to Torgerson, and 'it just became policy': *"As a result, we still don't know whether or not phonics works as the main teaching strategy for all children,"* she has said. *"Some of the recent evaluation work has demonstrated synthetic phonics may not be having the impact that was hoped for. If we'd done a randomised trial we would have known before the policy went national."*<sup>80</sup>

The *Magenta Book* states that there is direct link between quality and research design when thinking of causal questions.<sup>81</sup> High quality impact evaluation is all about the issue of attribution – i.e. can you attribute your intervention as the cause of the outcomes? – and to do this you need a good comparison or control group. Then you can be more confident that your particular good idea improved, say, children's services, reduced crime, or helped jobseekers back to work.

We should flag here that there are heated and expansive debates about what constitutes causality and attribution.<sup>82</sup> Many researchers would dispute the privileged position of experimental designs because they are impractical and don't always tell you why something that works in one place can work in other places.<sup>83</sup>

This definition of quality may not just be disputed by specialist researchers alone. For anybody outside evaluation, it may seem rather odd to equate quality with experimental designs. It's perfectly possible to have high quality research in other approaches, such as ethnography, focus groups or online surveys. It's also just as easy to have a shoddy RCT as in any other type of research.

### High quality 'qual' research

Many have sung the praises of good quality 'qualitative'<sup>84</sup> research for policymakers (see page 78 on Sunderland City Council's use of ethnography). The well-seasoned evaluation expert Stephen Morris at the independent social research institute NatCen says the most persuasive evidence is 'rigorous, social science'. He argues that we should not privilege quantitative over qualitative evidence, or vice versa. Qualitative evidence can be just as scientifically credible as quantitative.<sup>85</sup> But what we should do is:

*"...privilege evidence that involves the collection and analysis of data, through the application of defined, systematic and replicable methodologies, producing transparent data, testing predefined evaluation questions, and with results made publically available."*

(Evaluating service transformation, NatCen blog, 19 November 2014).

## Using top-tiers, hierarchies and standards of evidence

Another way of approaching this topic of quality is to use formal standards of evidence – particularly when looking at questions of ‘what works’ in social policy and practice.

These can take the form of hierarchies, where stronger research designs such as RCTs and systematic reviews are at the top for ‘causal’ questions – and other methods, usually case studies, are at the bottom.<sup>86</sup>

Figure C2: Two illustrations of simplified hierarchies of evidence, based on the design of the study

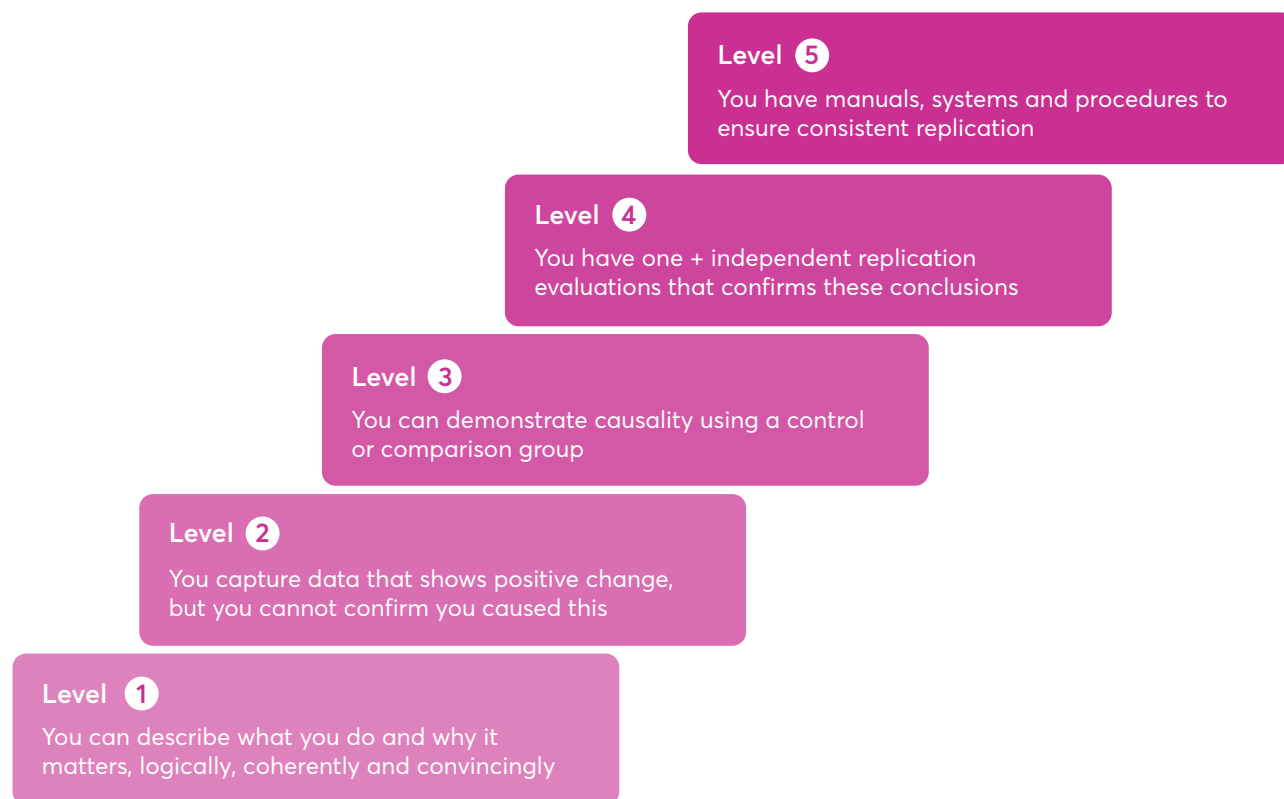


Source: Bagshaw and Bellomo 2008, p2.

Source: Petticrew and Roberts 2003, p.527.

Different versions of evidence hierarchies are used by many evidence review groups and endorsing bodies around the world, and they are particularly prevalent in healthcare. These standards of evidence have also been growing in UK social policy. Bodies that have been using them include Public Health England,<sup>87</sup> the Nesta/Cabinet Office's Centre for Social Action<sup>88</sup> and the youth crime reduction Project Oracle in London.<sup>89</sup>

Figure C3: Nesta's Standards of Evidence



Nesta developed its Standards of Evidence framework in 2012 to assess the quality of evidence on the impact of the social ventures we have invested in. Impact investment is big business and it has been given considerable support by the UK government. But for Nesta's investment fund, the concern was around how we could be sure that the new innovations were really having a positive impact. Could they even be doing harm? Strong evidence was needed to show if real progress was being made with the investments.

The Nesta framework has five levels of evidence.<sup>90</sup> These start from the basic level (level 1) where you can describe what you do and why it matters, logically, coherently and convincingly. You then move up the scale towards routine data collection (level 2), and then to higher levels where you use comparison groups (level 3), before creating evidence of replication and scaling (levels 4 and 5). The level of evidence you require should be appropriate for where you are in developing a policy or intervention. For Nesta, it's been particularly helpful for the ventures we fund to articulate how their product or service leads to positive social change.<sup>91</sup>

Critics of these hierarchies say that they can be overly-rigid straightjackets. Decisions of quality need to change according to context and it shouldn't be one-size-fits-all. Interestingly, the National Institute for Health and Care Excellence (NICE) – one of the UK Government's What Works centres (see page 98) – has now dropped its hierarchy of evidence. Their former Chairman, Michael Rawlins has been outwardly critical of 'slavishly' following evidence hierarchies.<sup>92</sup> He was in favour of a more nuanced sense of deciding what is appropriate, rather than RCTs and systematic reviews always trumping the rest.<sup>93</sup>

Whatever the merits of these criticisms,<sup>94</sup> these hierarchies, frameworks and principles do at least provide a structure to check the evidence claims. The Nesta model has been tried-and-tested. They provide a relatively easy-to-grasp, non-technical structure for the 'evidence journey' that many organisations need to go through; from the basics of a 'theory of change' to multiple replication studies. Evidence standards offer a way of vetting the strength of effectiveness claims, and of avoiding thinking that all evidence is equal.

These standards are also proportionate and take into account early-stage innovations. A new project does not have to show its impact straight away – or commission an expensive evaluation from day one. New ideas need time to grow. But reaching the higher levels of evidence (levels 4 and 5) does beg the question of how you judge the strength of the body of evidence, not just a single study – something we will look at next with systematic reviews.

### Avoid cherry picking by using systematic reviews

It's important for decision-makers to think about the quality of the evidence 'base' – not single pieces of evidence, but aggregated collections of research.

Much of the thinking on quality set out above has focused on single studies in 'primary research' – in other words, original research such as experiments, surveys or suchlike. We must also think about 'secondary research', and the summarising and pooling together of existing research – a task that is likely to be desk-based. We are much more likely to find a charity leader or government official undertaking this sort of secondary work.

A common pitfall is to carry out a 'literature review'. These are highly dubious. As Ben Goldacre put it, they encourage the temptation to cherry pick – consciously or unconsciously – your preferred bits of evidence to fit with your preferred conclusions:

*"Instead of just mooching through the research literature, consciously or unconsciously picking out papers here and there that support [our] pre-existing beliefs, [we] take a scientific, systematic approach to the very process of looking for scientific evidence, ensuring that [our] evidence is as complete and representative as possible of all the research that has ever been done."*

Ben Goldacre (2012) *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients*, (London: Fourth Estate).

We've seen in the earlier section on cognitive biases that we can unconsciously fall into the trap of looking for evidence that fits our beliefs. So we need to be very careful in collecting all the research that is out there, even the research with inconvenient facts.

One way to avoid this cherry picking is to use what are called systematic reviews.<sup>95</sup> These approaches aim to be exhaustive, and to find as much of the research relevant to your question as possible. They use explicit methods to identify what can reliably be said on the basis of these studies. The reviews can screen studies for quality – on the basis of the research design and methods.



Systematic approaches also have the value of being explicit about how they searched for research – so, in theory at least,<sup>96</sup> others could replicate the systematic review, unlike the 'mooching around' described by Ben Goldacre above.

## Characteristics of a systematic review

- Has a clearly stated set of objectives with pre-defined eligibility criteria for studies;
- Has an explicit, reproducible methodology;
- Uses a systematic search that attempts to identify all studies that meet the eligibility criteria;
- Includes a formal assessment of the validity of the findings of the included studies;
- Produces a systematic presentation, and synthesis, of the characteristics and findings of the included studies.

Source: HM Treasury (2011) *The Magenta Book; Guidance for Evaluation*.

## Meta-analysis

Meta-analysis is often an important part of systematic reviews and usually involves bringing together a large amount of research findings – but using statistical analysis. Frequently, it means pooling the average effect sizes estimated in each individual research study.

Meta-analysis is perhaps best known for combining the results of randomised controlled trials, but it is also commonly undertaken on non-randomised data – from primary studies that use case-control to cross-sectional and cohort designs (see Table C1 on page 80).

## Distinguishing between meta-analysis vs meta-evaluation vs systematic reviews

There is an important distinction between systematic reviews and meta-analysis, and they should not be conflated. A systematic review uses systematic methods for reviewing literature. It can be considered a 'meta level' (about or beyond) process as it is research on research. The term 'meta-analysis' could therefore be used to describe all systematic reviews, but in practice the term is used only for statistical meta-analysis, where statistics are used to synthesise the numerical data from different studies.

There is also the term 'meta-evaluation' (Scriven 1969) which can mean 'evaluation of evaluations', and can include the synthesis of evaluation studies, but also has broader meanings (Scriven 1969, Gough *et al.* 2012b).<sup>97</sup> For a longer description of 'meta-evaluation' and how it is different to other types of reviews, see the *Magenta Book*.

## Reviews in a hurry

But systematic reviews and meta-analysis can be time-consuming, which is no good for urgent decision-making. You probably can't wait around for six months to complete a review in time for a big speech tomorrow, a funding bid deadline or a memo on a new policy for a minister. However, many of these systematic reviews are already 'in the bag', so you don't need to start from scratch. They are freely available in online libraries such as the ones curated by the Cochrane and Campbell Collaborations, discussed in Section D on where to go for evidence.

Nevertheless, you may be frustrated and fail to find what you are looking for on those websites. So your best bet is to commission a 'pared-down version of the systematic review'<sup>98</sup> such as a rapid evidence assessment (REA). REAs aim to review evidence in a hurry, without unduly compromising quality (see Section D for more on REAs).<sup>99</sup>

However, since the questions we want answered now are very likely to recur in years to come, at the institutional level it makes sense to commission high quality reviews that are regularly updated and easily accessible, which is precisely the role of the online libraries covered in Section D.

Note, however, that more formal ways of bringing together research don't always equate to quality. Judging the quality of your evidence review is still hard to do. The quality of your synthesis is only really as good as the quality of the studies it is based on.

## The importance of repetition and corroboration

There are other things that need to be taken into account when vetting the quality of a research summary. For instance, how many studies need to be included for you to be comfortable that a strong body of evidence exists? Perhaps only a couple of studies if they are really good? Or maybe dozens, or hundreds? A recent 'Evidence Check' of the Department for Education<sup>100</sup> mentions 12 Ofsted reports on primary schools that purport to show strong performance related to phonics teaching. But there are around 16,000 state-funded primaries in England, so 12 doesn't sound very much. And that's also avoiding the other serious question of whether we can really say that Ofsted inspections count as strong evidence of impact.

The reality is that there is no magic number of studies. Yet we can't ignore the fact that the size of the body of evidence is important: there is strength in numbers, and we must have repetition and corroboration. Even studies that have won many accolades need to be repeated, again and again. Amgen, a Californian drug company, tried to replicate 53 landmark cancer studies. The work, published in *Nature*,<sup>101</sup> only replicated six (11 per cent) of these important studies. This is deeply troubling as the studies have influenced drug companies and cancer treatments globally.

## 'Reviews of reviews' – how to judge high quality bodies of evidence

We should certainly be very wary of single studies – about new wonder drugs for Alzheimer's splashed on the front page of the *Daily Mail*, for example – that are never replicated.<sup>102</sup> But the size of the evidence is not the only thing to consider. A *How to Note* on judging the strength of evidence produced by the DFID lists four things to consider when checking bodies of evidence:

1. The (technical) quality of the studies constituting the body of evidence (or the degree to which risk of bias has been addressed);
2. The size of the body of evidence;
3. The context in which the evidence is set;
4. The consistency of the findings produced by studies constituting the body of evidence.

Source: DFID (2014) *How to Note: Assessing the Strength of Evidence*

Judging all these criteria is always going to be rather subjective – and relevant to the context of the policy question. It's really hard to give blanket advice for the right body of evidence for any policy. Another interesting approach is to bring together only collections of systematic reviews, and not to sift through just single studies – in other words, a 'review of reviews'. Of course, this may only work if systematic reviews already exist – and in some sectors there are few. A good model is the Education Endowment Foundation – an independent charity and 'What Works Centre' funded by the Department for Education and Sutton Trust, which is discussed in Section D.

For a list of other in-depth guides to research quality, see Appendix 3.

### Key messages from Section C Part 2

- To find evidence that you can trust **look for peer-reviewed research**. But note that peer review is far from perfect. There can be unconscious biases such as 'herding', or publication bias towards positive results, or even academic fraud.
- When looking at questions of impact and 'what works', use the **frameworks and formal standards of evidence** such as those used by Nesta, Project Oracle and others.
- One study is never enough in social policy. **Avoid making decisions based on single studies, and look for multiple replications**. There is strength in numbers, and we must have repetition and corroboration. Even studies that have won many accolades need to be repeated, again and again.
- Don't do a 'literature review' of research, as they encourage cherry picking – consciously or unconsciously – of evidence. **Use systematic reviews**, which aim to be exhaustive and can screen studies for quality – usually on the basis of the research design and methods.
- If you haven't got time to do a systematic review, or can't find a completed one that answers your question, commission a **rapid evidence assessment**.

# Section D

## Where should you look for evidence?

This section offers advice and resources to help you find the right evidence to support your case.

The screenshot shows a Guardian article titled "Greg Hunt uses Wikipedia research to dismiss climate change-bushfires link". The article is by Helen Davidson, dated Thursday 24 October 2013 04:45 BST. The article text states: "Environment minister warns against 'politicising the science', saying bushfires are simply part of 'the Australian experience'". A photograph of Greg Hunt is shown. A hashtag "#GregHuntResearch" is visible on the right side of the article. The article has 6,500 shares and 322 comments. The Guardian logo and navigation menu are visible at the top.

It was refreshingly honest of former Australian environment Minister Greg Hunt to admit that he got his evidence from surfing the net. *"I looked up what Wikipedia says,"* was his defence on the BBC World Service when quizzed about his government's views on bush fires. Mr Hunt's comments went viral. He was ridiculed on Twitter, where his statement spawned the hashtag #GregHuntResearch. To rub home the point, some mischievous Wikipedians updated his personal Wikipedia page with a note that he *"was quoted as saying he uses Wikipedia for important policy research"*.

You would hope that other custodians of policy would look for more robust research than Wikipedia, but the Australian politician is not alone. A recent survey<sup>103</sup> by Carnegie UK Trust found that most policymakers accessed evidence through the internet, newspapers and broadcast news.

We are missing a trick here. At the very least, Greg Hunt could have surfed the ubiquitous **Google Scholar** search engine. There is a wide range of trusted online research resources, many of them free and easy to access. This chapter will signpost just a few that we think will be useful to any policymaker, charity leader or frontline professional.

Figure d1: The UK Evidence Ecosystem for social policy



### The UK What Works centres in social (and local economic) policy

The launch of six ‘what works’ centres in 2013 has transformed the landscape of evidence for social policy and practice in the UK.<sup>104</sup> Their aim is to improve the way government and other organisations create, share and use quality evidence and they are overseen by the What Works Network run out of the Cabinet Office, as well as the Scottish and Welsh governments. There are now nine centres – seen in the orange central ‘sausage’ in the figure above – covering areas like crime, education and wellbeing. They are funded by the Economic and Social Research Council, the Big Lottery Fund and a whole range of government departments – including the Scottish and Welsh governments.

No longer can we use the excuse that research is too difficult to access because it’s hidden behind academic paywalls, or in some indecipherable jargon. Whether you are in a charity, central government or local authority, they provide accessible and actionable syntheses of research and evidence. A particular priority is to make sure their summaries of research are communicated in a way that resonates with decision-makers – not only policymakers, but

also commissioners of services or frontline practitioners like teachers and police officers. It's still early days for some of these What Works Centres, so not all of their resources are finished yet. But, as we publish, the following are useful and open resources that summarise research:

- **The Sutton Trust/Education Endowment Foundation's (EEF) Teaching and Learning Toolkit and EEF Early Years Toolkit** – evidence of impact on pupil attainment for interventions in the classroom or around the school, such as mentoring or sports participation.

**Audience:** teachers, parents, heads, early years professionals, governors but also any charity, local authority or policymaker.

- **What Works Centre for Local Economic Growth** – systematic reviews on local economic growth policy issues such as access to finance, apprenticeships, broadband, employment training, enterprise zones and transport.

**Audience:** anyone involved in making policy decisions that are targeting economic growth – especially local authorities, LEPs, government, and businesses.

- **What Works Centre for Crime Reduction Toolkit at the College of Policing** – over 300 systematic reviews, covering 60 different crime reduction interventions, have been identified and will be added to the online tool over time, such as CCTV, street lighting and prison visits to deter young offenders.

**Audience:** police and crime reduction professionals, policymakers, commissioners of public services, charities and voluntary groups with interest in crime reduction programmes and interventions.

The screenshot shows the 'Crime Reduction Toolkit' website. The navigation bar includes 'ABOUT US', 'RESEARCH', 'CRIME REDUCTION TOOLKIT', 'RESEARCH MAP', and 'GET INVOLVED'. The main heading is 'Crime Reduction Toolkit'. On the left, there are sections for 'Our effect scale', 'Our quality scale', and 'About the Crime Reduction Toolkit'. Below these is a 'Key' section for 'Quality of evidence' with five levels: No information, Limited quality, Moderate quality, Strong quality, and Very strong quality. A 'Filters' section for 'Impact on Crime' is also present. The main table lists interventions with columns for 'Intervention', 'Impact on crime', 'How it works', 'Where it works', 'How to do it', and 'What it costs'. The interventions listed are Alcohol ignition interlock, Alcohol tax and price policies, NEW Alternative education programmes, CCTV, and Cognitive Behavioural Therapy (CBT).

Intervention	Impact on crime	How it works	Where it works	How to do it	What it costs
	Effect	Mechanism	Moderator	Implementation	Economic cost
Alcohol ignition interlock	✓✓	⚙️	📍	❓	£
Alcohol tax and price policies	✓✓	⚙️	📍	❓	£
<b>NEW</b> Alternative education programmes	X✓	⚙️	📍	❓	£
CCTV	✓✓	⚙️	📍	❓	£
Cognitive Behavioural Therapy (CBT)	✓✓	⚙️	📍	❓	£



- **Early Intervention Foundation Early Intervention Programmes Library and Guidebook** – covers the evidence for over 50 early intervention programmes. It looks to encourage positive development and tackle a range of problems, including abuse and neglect, substance abuse and mental health problems. The guidebook is designed to provide advice for professionals developing early intervention programmes and systems.

**Audience:** commissioners of services, policymakers and practitioners across social policy.

- **Public Policy Institute for Wales' evidence reviews** – the PPIW provides short turnaround analysis and advice for Welsh government ministers. It also undertakes work on some of the key strategic challenges facing the Welsh government, such as tackling poverty. Its website features links to dozens of freely available completed research summaries including rapid evidence reviews, expert advice and analysis by PPIW, ranging from strategic transport planning to the impact of indebtedness in Wales.

**Audience:** Welsh government ministers, others with interest in Welsh social policymaking.

Keep an eye out for forthcoming online summaries of research by the other What Works Centres:

- What Works in Tackling Poverty (led by the Public Policy Institute for Wales – who also run the What Works Centre for Wales).
- What Works Scotland Evidence Bank.
- What Works Centre for Wellbeing.
- Centre for Ageing Better.

## NICE work in health and social care

A lodestar in the UK What Works centres is the National Institute for Health and Care Excellence (NICE). The guidance offered by NICE is all based on the best available research evidence – as well as inputs from patients through 'citizen jury' type bodies, and clinicians.<sup>105</sup> NICE also offers two other useful resources:

- **NHS Evidence:** a helpful database that gives digestible research and guidance ready-made for the doctor's surgery or hospital ward. It's an extraordinary resource and ranges from open and free databases on medicines to more advanced (but unfortunately only for NHS staff) links to top academic journals and bibliographic databases.
- **UK DUETS:** short for the Database of Uncertainties about the Effects of Treatments, it flags up where serious question marks remain about whether treatments work or not. It's based on what patients and clinicians are saying, as well as the best available primary research and systematic reviews. The database is candid about our gaps in understanding and recognises that 'ignoring treatment uncertainties puts patients at risk'.

## Other research databases for social policy and practice

- Most databases are not as practitioner-friendly as NHS Evidence, and are aimed more at specialist researchers. Some of these other databases can, however, be surprisingly easy to access:



- The **Evidence for Policy and Practice Information (EPPI)** Co-ordinating Centre at the University College London has a range of databases that are free and online, such as the **Database of Education Research**, which has 1,200 records of systematic reviews. You can search all its systematic reviews by keyword.
- There are also large scientific databases of journals, such as **PubMed** for studies in public health or **JSTOR** for studies in economics, sociology and public policy – although this is aimed mostly at academics. JSTOR is a 'not-for-profit', but it does seek a charge to cover its costs and it is mostly used in universities, public libraries and schools. If you work in the UK central government or local authority, your department may well subscribe and 'walk-in' users can access over 2,000 academic journals.
- If you still have close links to your old UK university, don't forget that alumnae often have remote access to their former university library databases, such as **Oxford University's Alumni** resource.

## Other systematic reviews beyond the What Works centres

In the previous section we recommend using systematic reviews, not just single studies. Although expensive to commission these reviews from scratch, there are many completed reviews that are relatively easy to access.

### General social policy

The EPPI-Centre at University College London mentioned above has an 'evidence library' of systematic reviews that anybody can access, although the range is not comprehensive as it only covers those studies that the Centre has been involved in.

A more comprehensive website for lists of systematic reviews in social policy is the Campbell Collaboration Library. It covers education, crime and justice, social welfare and international development. Managed by a Secretariat, Campbell is organised into Coordinating Groups who manage a rigorous peer review process for high quality systematic reviews. The reviews can be on highly topical subjects that are popular with decision-makers, such as its reviews of 'mindfulness' in stress reduction that has had 10,000 downloads.

### Education

Another alternative free resource to the Education Endowment Foundation What Works Centre is the Evidence for Impact website that provides a simple evidence rating system for programmes relating to schools, along with concise evidence summaries. It has a comprehensive database of programmes available in the UK, including details on their effectiveness and cost, together with links to the providers and experts who can offer further support. The database can be easily searched by key stage, subject area and targeted group, so that results are tailored to the specific needs of your class or school.

### Children's services

Dartington Social Research Unit has a one-stop-shop for children's service commissioners called Investing in Children. It brings together evidence on 'what works' with the economic costs and benefits of 100 different competing interventions in children's services.

### International development

For international development there is the 3ie database covering both primary studies and systematic reviews of the effectiveness of social and economic interventions in low- and middle- income countries.

Rather than go through each of the above portals or websites, you may consider going straight to Google Scholar. As a tip, to find systematic reviews use the search terms 'subject' AND 'systematic review' e.g. 'mindfulness' AND 'systematic review'. If you want reviews from specific organisations you can expand the search string to specify that, e.g. 'mindfulness' AND 'systematic review' AND 'Campbell' or another organisation that produces reviews.

## Last stop is a one-stop-shop

There are multiple one-stop-shops that collate evidence resources around a particular topic, but the danger is that they fall prey to 'Portal Proliferation Syndrome'. In other words, producers of research can't resist the temptation to set up another hub, portal, platform or gateway to try and package information all in one place. New databases of databases may only confuse users.

Still, pooling information into a ready-to-use form for social policy professions is surely a noble cause. And bringing together databases in one place can help reduce confusion if done well. There are some great synthesising digital resources out there. The Social Policy and Practice (SPP) one-stop-shop for research in health and social policy responded to the proliferation of information resources by splicing together four previous national research databases relevant to policy: ChildData, AgeInfo, Planex and Social Care Online.

SPP is useful for any professional working in the field of social care or social work who can't get easy access to a university library. As well as helping to access 600 UK and international journals, it also includes the 'grey literature' of government or think-tank reports and guidance that will help those working in social care. You can find a brief guide to these UK research databases on the Alliance for Useful Evidence website.<sup>106</sup>

## Short, snappy bitesize summaries

Some Campbell Collaboration reviews can be up to one hundred pages long, but a time-poor decision-maker will need something shorter. Abstracts may tell you a fair amount, and there are plenty of academic journals that provide their abstracts – if not full articles – for free. Campbell reviews contain a short 'plain language summary', which are currently being made more accessible.

There are other readable and non-academic narratives that summarise reviews of research, written in a way that is practitioner-friendly (i.e. dropping academic jargon unless essential), such as the 2011 Scottish Government's summaries of evidence<sup>107</sup> on what works in stopping people from reoffending, followed by the 2013 UK Government's summary of evidence<sup>108</sup> on the same topic. Or the topical articles by academics on The Conversation website, including the Hard Evidence section.

Parliamentary POSTNotes are also highly digestible short PDF research summaries that anybody can access, not just MPs. Although they mostly cover science and technology, they have a growing number in social science and social policy. You should also subscribe to the Economic and Social Research Council regular Evidence Briefings; and if your interest is in children's policy, you can still find Evidence Nuggets from the What Works Centre for Children (now closed).<sup>109</sup>

## Consultancy and helpdesk-type services – paying the price for evidence

However, some research one-stop-shops are not open to everybody or may require a subscription or payment. But if the product is good, isn't it worth the investment? Some of us may be lucky enough to have access to university libraries (and those unsung heroes of evidence, good librarians) so that you can get behind the paywalls of prestigious research journals. But for those of us that cannot (including the authors of this guide, despite working for a large UK charity), is it worth paying somebody to track down your evidence?

One example of a fee-charging provider is the Idox Information Service. It has been around for 30 years now and claims to have over 200,000 users. As well as having a database of abstracts, there are consultants on hand to help you. It even has a good old fashioned library service – you can rent books out by post. Idox claim to be the 'most complete and accessible source of evidence and research for the public sector in the UK', but you have to pay for it.

There are other research consultancies that can help with locating evidence – for a list of these who are also members of the Alliance see the website.

One method that might be helpful is a rapid evidence assessment (REA). They aim to be fairly comprehensive, but have the downside of not being exhaustive, with a consequent risk of bias.<sup>110</sup>

This table from the National Foundation for Educational Research shows what you might get if you commissioned an REA, or other type of quicker review such as a scoping study:

Table D2:

Product	Why would I use it?	What will it give me?	How long will it take?	Suggested report length
<b>Rapid review</b>	To get a quick overview of an area of policy or practice	An overview of what we know about a specific topic. It can be based on research literature, policy or practice evidence. Literature is located through limited searching	1–2 months	10-page report
<b>Scoping review</b>	To find out if there is enough evidence to merit a rapid evidence assessment or full systematic review	A map of how much evidence there is on a topic	1–2 months	10-page report
<b>Rapid evidence assessment (REA)</b>	To get a thorough evidence synthesis to inform policy or practice. To answer questions about what is effective	A robust synthesis of the evidence, with key messages and recommendations. Literature is found through comprehensive searching	3–5 months	15-page report
<b>Meta-analytical review</b>	To pool results from a series of high quality quantitative evaluations	A numeric estimate of the overall effect of the intervention being evaluated	3–5 months	15-page report
<b>Full systematic review</b>	To get a 'gold standard' evidence synthesis of all available evidence	A highly robust and systematic synthesis of the evidence, with key messages and recommendations. Literature is found through highly comprehensive searching	6–12 months	60-page report

Source: The National Foundation for Educational Research, *Investigating what works: through reviews*

We have to tread carefully here. Endorsing commercial providers such as Idox (or potentially libelling them if we are critical) is difficult territory. But we felt we couldn't avoid mentioning evidence aggregators that demand a fee of some sort. And the reality is that one of the biggest conduits for research – academic journals – are not (yet) open access and require a fee. If we have missed any database resources that you think are useful, let us know so we can add them to our online resources.

## Expert advice

Sometimes, passive databases are not enough. We might not even be sure exactly what our question is – and therefore what research we should be looking for. This is when a conversation with an expert may be helpful, so it is reassuring to hear of a new service being developed in Scotland, the Evidence Bank, developed by the Centre for Research on Families and Relationships (CRFR). As well as providing evidence, it has also piloted a special request service. The model was developed in partnership with organisations and services in the third and public sectors, and covers children, families, relationships and disability. The Evidence Bank is currently being developed as part of What Works Scotland and should be online soon.

In Westminster and Whitehall, academics have also started an Evidence Information Service<sup>111</sup> but it's in its infancy so watch this space to see when it is fully up and running.

If you are in a major government department in Whitehall, Scotland, Wales or Northern Ireland, you are lucky to have in-house analysts, researchers, statisticians and economists to call upon to help you find – and interpret – evidence. For instance, there is the Government Social Research Service.

This chapter has outlined a variety of easily accessible UK research resources for social policy and practice. However, once you have gathered your evidence, you don't want it gathering dust. We need to think about how to communicate and then act on it. How best to do that is the focus of the next and final section.

## Key messages from Section D

- Beware of **haphazard online searches** – use trusted repositories of research.
- Take advantage of the **freely available, accessible and actionable summaries of research**, such as those on the websites of nine UK What Works Centres, NHS Evidence, EPPI Centre and Campbell and Cochrane Collaborations.
- Read some of the **free, bitesize digests of research** such as the UK Parliament's POSTNotes or Economic and Social Research Council regular Evidence Briefings.
- If you can't find what you are looking for, you should **consider commissioning a consultancy or helpdesk-type service** to do a research summary, such as a rapid evidence assessment, scoping review or systematic review.
- If you are in a government department or local authority, **use your in-house analysts** to do the search for you.

## Section E

# How should you communicate your findings?

This section focuses on how to get your message across once you have the evidence you need.

There's not much point in gathering a strong body of research if it falls on deaf ears. The whole point is to transfer it into practice and change attitudes, beliefs and behaviour.

Successful communication means really putting yourself in the shoes of your audience, with many official research uptake guides saying ideally this should start with the evidence gathering process itself.<sup>112</sup>

The best bet is to avoid passive dissemination, pushing information out through the door in the hope that some of it sticks. Instead, you need to actively engage with your audiences.

For instance, the Medical Research Council's (MRC) research guide<sup>113</sup> has some helpful advice on research uptake and features guidance based on published research (rather than expert opinion). It covers social policy as well as traditional medicine, and includes case studies in areas such as crime prevention and public health.

### Six tips from the Medical Research Council's guide on getting your evidence heard

1. **Involve stakeholders** in the choice of question and design of the research to ensure relevance.
2. Provide evidence in an **integrated and graded way**: have reviews, not individual studies, and variable length summaries that allow for rapid scanning.
3. **Take account of context**, and identify the elements relevant to decision-making, such as benefits, harms and costs.
4. **Make recommendations as specific** as possible.
5. Use a **multifaceted approach involving a mixture of interactive** – rather than didactic – educational meetings, feedback, reminders and local consensus processes.
6. Successful implementation depends on changing behaviour – often of a wide range of people. This requires a **scientific understanding of the behaviours** that need to change, the factors maintaining current behaviour and barriers and facilitators to change, and the expertise to develop strategies to achieve change based on this understanding. Further research may be needed to assist the process of implementation, and implementation research teams should include a behavioural scientist.

Source: Medical Research Council (2013) *Developing and evaluating complex interventions: new guidance*

Unfortunately, as the MRC guide alludes to, the evidence base for effective implementation remains limited. There's plenty of 'good practice guides' and advice, but most of it does not seem to be based on evidence – just what counts as expert opinion. And what evidence we do have appears to be inconsistent with a lot of current practice.

For instance, the obsession with short summaries for policymakers is based on shaky evidential grounds. A multi-arm RCT by the International Initiative for Impact Evaluation (3ie) and others<sup>114</sup> explored the effectiveness of a policy brief for influencing readers' beliefs and prompting them to act.

*"The results are striking, and not that reassuring for those, including 3ie, who place importance on policy briefs,"* then-Executive Director of 3ie, Howard White, has said.<sup>115</sup> For example, the trial found that changing the nature of the brief – such as its length or including authoritative op-eds – was irrelevant to changing beliefs. However, the 'messenger' of the briefing – i.e. who presents it – did matter to whether somebody might act on it. There is a considerable amount of evidence from social psychology on the 'messenger' that we will touch on below as it gives some helpful direction on better communication of policy.

## Guides on research uptake

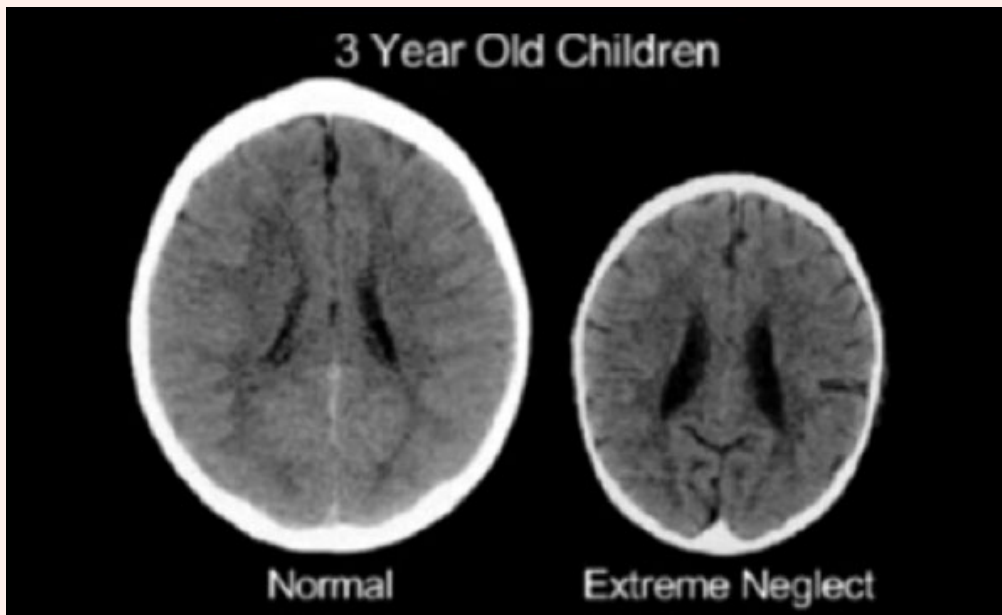
- **Research Uptake Guidance** – Department for International Development (UK)
- **Communicating research for evidence-based policymaking: A practical guide for researchers in socio-economic sciences and humanities** – European Commission
- **Guide: Engaging with Policymakers** – National Coordinating Centre for Public Engagement (UK)
- **Impact Toolkit** – Economic and Social Research Council (UK)
- **Helping researchers become policy entrepreneurs: How to develop engagement strategies for evidence-based policymaking** – Overseas Development Institute (UK)

## Make it Easy, Attractive, Social and Timely

If you want to encourage a behaviour change based on research, make it **Easy, Attractive, Social and Timely (EAST)**. These four simple principles are based on the Behavioural Insights Team's own work and wider academic literature:<sup>116</sup>

- **Easy** – e.g. **simplify the message**. This is probably one of the most common messages of communicating research to lay audiences. This may even be as simple as the format of the written briefing. Eye-tracking research conducted for Royal Mail in 2010<sup>117</sup> suggests that **people generally focus on headings, boxes and images, while detailed text is often ignored**. However, simplification of messages can be overstressed. There is some evidence that complexity has more lasting learning outcomes, if you are really trying to change people's knowledge and understanding.<sup>118</sup>
- **Attractive** – e.g. **use images or personalisation**. This helps to attract attention (see box on following page on use of images and MRI scans).
- **Social** – e.g. **use anecdotes, and real people**. Tap into networks to help encourage wider behaviour change, and encourage people to make a commitment to each other.
- **Timely** – we are more influenced by **immediate costs/benefits** rather than later ones (a challenge to early intervention) so make it current. Also get your timing right and prompt people when they are likely to be most receptive; the same offer made at different times can have drastically different levels of success.

## The power of images – and brain scans



Perry, B.D. (2002) Childhood experience and the expression of genetic potential: what childhood neglect tells us about nature and nurture. 'Brain and Mind'. 3: 79–100.

This image has been highly influential and was put on the front page of a well-evidenced and respected report for government by MP Graham Allen<sup>119</sup> making the case for early intervention. The right-hand brain appears to show the impact on the brain of a child suffering 'extreme neglect' and damage for their subsequent lives, perhaps based on bad parenting. Some, however, have queried the use of such MRI images.<sup>120</sup> What, for instance, was the case-history of the two children's brains above? Perhaps the child on the right had other severe disabilities.

But we do know how powerful neuro-imaging can be on policymakers – and even on other experts. One RCT experiment found that including brain images was linked to higher perceptions of scientific merit. University students were shown identical academic papers, the only difference being that one included brain images, one had none (and one with other images, like traditional bar graphs). People thought the paper with the brain image had much more scientific credibility, without realising that it was exactly the same text and research.<sup>121</sup>

### Key messages from Section E

- Successful communication means **empathising with your target audience** – ideally from the start of the evidence gathering process.
- Avoid passive dissemination. Instead, you need to **actively engage with your audiences**.
- If you do have to be more passive in your communication, use the Behavioural Insights Team's tips on behaviour change. **Make it Easy, Attractive, Social and Timely**.



# Summary

Whether you are a policymaker or leader in a charity or local authority, evidence is unlikely to be a nice 'fit' with all the other competing sources of information. Research evidence is – and should be – just part of the mix.

How to go about smarter decision-making generally is a whole new practice guide in itself. Only pointers and checklists can be given here. There will never be a magic formula for how to 'do' evidence-informed policy. There are, however, models and tools to help decision-making. For policymakers, **Harvard's Evidence for Policy Design** team has created online tools such as 'decision-trees' or a 'policy analysis matrix' to help weigh up the competing pieces of information behind a policy, thanks to funding from the UK's DFID.<sup>122</sup> These decision-making tools are helpful but will never replace professional judgment in the messy and complex world of social policy. They can be a prop or checklist to support it. Instead of our own list, we have two key recommendations for using evidence to inform better decision-making:

## 1. Implement changes or advice based on evidence that is as strong as possible (and vice versa, be wary of changes or advice if the evidence is weak)

The advice of this guide is that the stronger the evidence, particularly if it is based on not just one study but multiple replications and systematic reviews, the more you should be confident of it being a part of your decision.

What we shouldn't be doing is making bolder policy recommendations based on weaker evidence. That may seem obvious to you, but it doesn't seem to be happening in policymaking. The National Audit Office review of government evidence (see page 88) found that the weakest evaluations were more likely to make 'bold, un-caveated claims' about the positive impacts of the policy examined.<sup>123</sup> If there is weak – or non-existent – evidence, then be honest about it and say so. Over-claiming is a dangerous game. And although we know how frustrating it is to see a conclusion of 'we need more research' in policy reports or advice, it's also true: we will never stop needing to understand and research.

## 2. Be evidence-aware in your decision-making

The second piece of advice is to be self-aware about potential cognitive biases in your decision-making. Being alert to the fact that you may be leaping to the wrong conclusions – even when you have located the best available research – is paramount.

Finally, we want to stress that the single most important message of this guide is appropriateness: 'horses for courses'. You need to think about the right type of research to suit your needs. It has to be proportionate – a massive, multi-armed RCT would be madness for a small rural-based social enterprise. But likewise, a decade-long £50 million social programme should seriously invest in evidence.

The type of evidence also has to match your question. Does it fit where you are in growing a programme, policy or practice? To help you make that judgement, use some of the formal framework, such as the Nesta Standards of Evidence, if your question is about impact. Also, think about where you are in the Nesta innovation spiral – would learning from others be your best choice, or is it time to formally experiment on your growing idea?

And don't be put off by expense – we have given you some handy pointers to where research is already available, and for free. Research is becoming easier to source. Academics are under more pressure to step outside their university walls and help you in policymaking, charitable work or delivering good local services.

We also need to be smarter on how to communicate evidence. Evidence rarely speaks for itself. A common communication message is to simplify, visualise, empathise and be 'translational' – in other words, interact with your audience. In the coming months and years, we will be doing just that: creating videos and interactive digital media for blended learning, and perhaps, dear reader, meeting you face-to-face in one of our ongoing Evidence Masterclasses.

# Appendix 1

## Rationales for evidence and types of evidence required for policymakers

Source: Louise Shaxson (2014) *Investing in Evidence; Lessons from the UK Department for Environment, Food and Rural Affairs*.

Heading	'Big questions'	Rationales for evidence needs	Types of evidence required
A: <b>Understanding the context</b> ; fundamental processes and phenomena, baselines and benchmarks	Where are we now?	<ul style="list-style-type: none"> <li>• To gather and analyse available/new data</li> <li>• To evaluate risks, issues and uncertainties</li> </ul>	<ul style="list-style-type: none"> <li>• Reviews of existing knowledge</li> <li>• Surveys of social and environmental data</li> <li>• Research on causality</li> <li>• Risk assessment</li> </ul>
B: <b>Development</b> of models, methodologies and tools	Where are we going?	<ul style="list-style-type: none"> <li>• To understand current drivers and trends</li> <li>• To predict future drivers and trends</li> <li>• To assess implications for policy outcomes</li> </ul>	<ul style="list-style-type: none"> <li>• Sensitivity analysis</li> <li>• Horizon scanning</li> <li>• Forecasting and scenarios</li> <li>• Modelling impacts and outcomes</li> </ul>
C: Developing and using the evidence base to help <b>set targets and formulate policy</b>	Where do we want to be over the next 5-10 years?	<ul style="list-style-type: none"> <li>• To understand the economic/social value of change</li> <li>• To understand the feasibility/cost of change</li> <li>• To negotiate goals</li> </ul>	<ul style="list-style-type: none"> <li>• Economic and social research</li> <li>• Deliberative engagement processes</li> <li>• Feasibility and pilot studies</li> <li>• Market surveys</li> </ul>
D: Development and appraisal of <b>options/solutions</b>  E: Optimum decisions and effective implementation through <b>communication</b> , engagement and consultation to influence change	How do we get there?	<ul style="list-style-type: none"> <li>• To identify/evaluate current options</li> <li>• To identify/develop new solutions</li> <li>• To evaluate new/old options</li> </ul>	<ul style="list-style-type: none"> <li>• Option/evaluation studies</li> <li>• Regulatory impact assessments</li> <li>• Interventions to promote innovation</li> </ul>
F: <b>Monitoring progress</b> towards policy/programme targets  G: Policy/programme <b>evaluation</b>	How well did we do?	<ul style="list-style-type: none"> <li>• To monitor progress</li> <li>• To evaluate policies and programmes</li> <li>• To learn lessons</li> </ul>	<ul style="list-style-type: none"> <li>• Interdisciplinary evaluations</li> <li>• Deliberative evaluation processes</li> </ul>

# Appendix 2

## Experimental designs for evaluating complex interventions

This list of five different experimental designs is taken from *Developing and evaluating complex interventions: new guidance*, prepared on behalf of the Medical Research Council (2013).

### 1. Individually randomised trials

Individuals are randomly allocated to receive either an experimental intervention, or an alternative such as standard treatment, a placebo or remaining on a waiting list. Such trials are sometimes dismissed as inapplicable to complex interventions, but there are many variants of the basic method, and often solutions can be found to the technical and ethical problems associated with randomisation.

### 2. Cluster randomised trials

Contamination of the control group, leading to biased estimates of effect size, is often cited as a drawback of randomised trials of population level interventions, but cluster randomisation, widely used in health service research, is one solution. Here, groups such as patients in a GP practice or tenants in a housing scheme are randomly allocated to the experimental or a control intervention.

### 3. Stepped wedge designs

The randomised stepped wedge design may be used to overcome practical or ethical objections to experimentally evaluating an intervention for which there is some evidence of effectiveness, or which cannot be made available to the whole population at once. It allows a randomised controlled trial to be conducted without delaying roll-out of the intervention. Eventually, the whole population receives the intervention, but with randomisation built into the phasing of implementation.

### 4. Preference trials and randomised consent designs

Practical or ethical obstacles to randomisation can sometimes be overcome by the use of non-standard designs. Where patients have very strong preferences among treatments, basing treatment allocation on patients' preferences, or randomising patients before seeking consent, may be appropriate.

### 5. N-of-1 designs

Conventional trials aim to estimate the average effect of an intervention on a population, and provide little information about person variability in response to interventions, or about the mechanisms by which effective interventions achieve change. N-of-1 trials, in which individuals undergo interventions with the order or scheduling decided at random, can be used to assess between and within person change, and to investigate theoretically predicted mediators of that change.

# Appendix 3

## Guidance on evidence quality, designs and methods

Scottish Government Social Researchers' Method Guides – short introductions to some of the most common methods used in social science research  
[www.gov.scot/Topics/Research/About/Social-Research/Methods-Guides](http://www.gov.scot/Topics/Research/About/Social-Research/Methods-Guides)

The Bond Evidence Principles and checklist – specially designed for NGOs  
[www.bond.org.uk/effectiveness/monitoring-and-evaluation](http://www.bond.org.uk/effectiveness/monitoring-and-evaluation)

Evidence for Success: the guide to getting evidence and using it – a practical guide for third sector organisations from Evaluation Support Scotland and KTN  
[www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised\\_april\\_2015-web.pdf](http://www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised_april_2015-web.pdf)

Research Guide for Third Sector Organisations  
[www.ncb.org.uk/media/858179/research\\_guide\\_for\\_third\\_sector\\_organisations.pdf](http://www.ncb.org.uk/media/858179/research_guide_for_third_sector_organisations.pdf)

The Medical Research Council's Developing and evaluating complex interventions: new guidance – provides guidance on the development, evaluation and implementation of complex interventions to improve health  
[www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/](http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/)

Government Office for Science – The Government Chief Scientific Adviser's Guidelines on the Use of Scientific and Engineering Advice in Policymaking  
[www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf)

National Institute for Health and Care (NICE) – The Guidelines Manual: Reviewing the Evidence  
[www.nice.org.uk/article/pmg6/chapter/6-reviewing-the-evidence](http://www.nice.org.uk/article/pmg6/chapter/6-reviewing-the-evidence)

DFID – How to Note: Assessing the Strength of Evidence  
[www.gov.uk/government/publications/how-to-note-assessing-the-strength-of-evidence](http://www.gov.uk/government/publications/how-to-note-assessing-the-strength-of-evidence)

HM Treasury – The *Magenta Book*: Guidance for Evaluation  
[www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/220542/magenta\\_book\\_combined.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf)

Civil Service Learning – Policy Profession: Skills and Knowledge Framework  
[civilservicelearning.civilservice.gov.uk/sites/default/files/policy\\_profession\\_skills\\_and\\_knowledge\\_framework\\_jan2013web.pdf](http://civilservicelearning.civilservice.gov.uk/sites/default/files/policy_profession_skills_and_knowledge_framework_jan2013web.pdf)

HM Treasury, DECC and DEFRA – Quality in policy impact evaluation; understanding the effects of policy from other influences (Supplement to Magenta Guide)  
[www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/190984/Magenta\\_Book\\_quality\\_in\\_policy\\_impact\\_evaluation\\_QPIE\\_.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation_QPIE_.pdf)

Research Councils, Universities UK *et al.* – UK concordat to support research integrity  
[www.universitiesuk.ac.uk/highereducation/Pages/Theconcordattosupportresearchintegrity.aspx#.Vd2X5fVhBc](http://www.universitiesuk.ac.uk/highereducation/Pages/Theconcordattosupportresearchintegrity.aspx#.Vd2X5fVhBc)

### Guides to commissioning research

Social Research Association – Commissioning Social Research good practice guide  
[the-sra.org.uk/sra\\_resources/research-commissioning/](http://the-sra.org.uk/sra_resources/research-commissioning/)

Market Research Society guide to 'buying research'  
[www.mrs.org.uk/intelligence/research\\_buyers\\_guide](http://www.mrs.org.uk/intelligence/research_buyers_guide)

# Endnotes

1. The Police Foundation (2013) Policing and Crime Reduction: The evidence and its implications for practice. See page 11: [http://www.police-foundation.org.uk/uploads/holding/projects/policing\\_and\\_crime\\_reduction.pdf](http://www.police-foundation.org.uk/uploads/holding/projects/policing_and_crime_reduction.pdf)
2. Sense about Science (2015) Making Sense of Crime: Evidence on the causes of crime and policies to reduce it. See page 17.
3. OECD (2015) Students, Computers and Learning: Making the Connection. See: <http://www.oecd.org/publications/students-computers-and-learning-9789264239555-en.htm>
4. For an overview of evidence on 'what works' in the classroom for pupil attainment, see the Education Endowment Foundation Toolkit: <https://educationendowmentfoundation.org.uk/toolkit/toolkit-a-z/>
5. This is taken from the 'Policy Profession Skills and Knowledge Framework' (see endnote 9) but it is just as relevant to our audience of this guide – working outside policy, in local authorities and charities.
6. Sackett, D., Rosenberg, W., Gray, M., Haynes, B. and Richardson, S. (1996) Evidence based medicine: what it is and what it isn't. 'British Medical Journal'. 312:71. See: <http://www.bmj.com/content/312/7023/71>
7. The definition came at a time when most medical decision-making was based on experience, authority and eminence. Medical practice was not informed by the best available scientific evidence. Commentators and researchers like Ben Goldacre have argued that social policy is in the same place as medicine 20 years ago: authority still reigns, not research.
8. Graber, M.L., Franklin, N. and Gordon, R. (2005) Diagnostic Error in Internal Medicine. 'Internal Medicine.' 165(13). 1493-1499. See: <http://archinte.jamanetwork.com/article.aspx?articleid=486642&resultclick=1>
9. Civil Service Learning (2013) 'Policy Profession Skills and Knowledge Framework.' London: Civil Service Learning. See: [https://civilservicelearning.civilservice.gov.uk/sites/default/files/policy\\_profession\\_skills\\_and\\_knowledge\\_framework\\_jan2013web.pdf](https://civilservicelearning.civilservice.gov.uk/sites/default/files/policy_profession_skills_and_knowledge_framework_jan2013web.pdf)
10. <http://www.oxforddictionaries.com/definition/english/evidence>
11. Frost, S. et al. (2006) 'The Evidence Guide: Using Research and Evaluation in Social Care and Allied Professions.' London: Barnardo's
12. Nutley, S., Powell, A. and Davies, H. (2013) 'What Counts as Good Evidence?' London: Alliance for Useful Evidence
13. Adapted from Shaw, I. (1999) 'Qualitative Evaluation.' London: Sage.
14. Tetlock, P.E. (2005) Expert Political Judgement: How Good Is It? How Can We Know?. New Jersey: Princeton University Press
15. Vermeren, P. (2015) 'Why the CEBMa model is misleading.' See: [http://evidencebasedhrm.be/wp-content/uploads/2015/08/Why-the-Cebma-model-is-misleading\\_august\\_2015.pdf](http://evidencebasedhrm.be/wp-content/uploads/2015/08/Why-the-Cebma-model-is-misleading_august_2015.pdf)
16. Ariely, D. (2009) Predictably Irrational: The Hidden Forces that Shape Our Decisions. London: HarperCollins.
17. See also White, H. and Phillip, D. (2012) 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework.' – a working paper from the International Initiative for Impact Evaluation that discusses how evaluators themselves can be very biased: "Bias affects all evaluation approaches and methods; it affects the choices that evaluators make, the questions that they ask, the data they collect and the way that they carry out analysis" (page 24).
18. The chances of social workers getting it wrong are compounded by the sheer burden on their decision-making – they have to make many tough sequential decisions through the course of a single day, which engenders 'decision fatigue'. For a discussion of this research see page 7 of Breckon, J. and Hay, J. (2015) 'Knowing How to Protect; Using Research Evidence to Protect Children from Harm'. London: Alliance for Useful Evidence/NSPCC. See: <http://www.alliance4usefulevidence.org/assets/NSPCC-report-FINAL-V-for-printers.pdf>
19. Kassina, S.M., Dror, I.E. and Kukuckaa, J. (2013) The forensic confirmation bias: Problems, perspectives, and proposed solutions. 'Journal of Applied Research in Memory and Cognition.' 2(1). 42-52. See: [http://cognitiveconsultantsinternational.com/Dror\\_JARMAC\\_forensic\\_bias.pdf](http://cognitiveconsultantsinternational.com/Dror_JARMAC_forensic_bias.pdf)
20. Kahneman, D., Lovallo, D. and Sibony, O. (2012) 'The Big Idea: Before You Make That Big Decision...'. See: <https://hbr.org/2011/06/the-big-idea-before-you-make-that-big-decision>
21. Flyvbjerg, B., Skamris Holm, M.K. and Buhl, S.L. (2002) Underestimating Costs in Public Works Projects – Error or Lie. 'Journal of the American Planning Association.' 68(3). 279-295.
22. See page 7 of Lewis, P.G. (2013) 'Policy Thinking, Fast and Slow: A Social Intuitionist Perspective on Public Policy Processes'. APSA 2013 Annual Meeting Paper, American Political Science Association 2013 Annual Meeting. See: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2300479](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2300479)
23. See Cairney, P. (2014) 'The Psychology of Evidence Based Policymaking: Who Will Speak for the Evidence if it Doesn't Speak for Itself?' See: <https://paulcairney.wordpress.com/2014/07/1288/>
24. Kahneman D. and Klein G. (2009) Conditions for intuitive expertise: a failure to disagree. 'American Psychology.' 64(6). 515-26. See: <http://www.ncbi.nlm.nih.gov/pubmed/19739881>
25. Gornall, J. (2008) Does cot death still exist? 'British Medical Journal.' 336(7639). 302-304. See: <http://dx.doi.org/10.1136/bmj.39455.496146.AD>
26. Nesta (2013) 'Understand how innovation works.' Video available at <http://www.nesta.org.uk/resources/understand-how-innovation-works>
27. Gabriel, M. (2014) 'Making it Big: Strategies for scaling social innovations.' London: Nesta. See: <http://www.nesta.org.uk/publications/making-it-big-strategies-scaling-social-innovations>
28. White, H. (2015) 'Learning what works and why from evidence of what doesn't work.' See: <http://www.alliance4usefulevidence.org/learning-what-works-and-why-from-evidence-of-what-doesnt-work/>
29. Nesta and TSIP (2014) 'Guidance for Developing a Theory of Change for Your Programme.' See: [https://www.nesta.org.uk/sites/default/files/theory\\_of\\_change\\_guidance\\_for\\_applicants\\_.pdf](https://www.nesta.org.uk/sites/default/files/theory_of_change_guidance_for_applicants_.pdf)
30. See page 3 of Stern, E. (2015) 'Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.' London: Big Lottery Fund, Bond, Comic Relief and the Department for International Development.
31. Johnson, S. (2011) 'Impact of social science on policy: Sure Start case study.' Report to Economic and Social Research Council (ESRC)
32. Mulgan, G. (2015) 'The six Ws: a formula for what works.' London: Nesta. See: <http://www.nesta.org.uk/blog/six-ws-formula-what-works>
33. Nesta (2013) 'What is ethnographic research?' Video available at <https://vimeo.com/77509484>
34. Petticrew, M. and Roberts, H. (2003) Evidence, hierarchies and typologies: horses for courses. 'Journal of Epidemiology and Community Health.' 57. 527-529.
35. For an authoritative and exhaustive list of social science research frameworks and methods see Luff, R., Byatt, D. and Martin, D. (2015) 'Review of the Typology of Research Methods within the Social Sciences.' London: ESRC/National Centre for Research Methods. This typology is one of the most downloaded parts of the NCRM website. See: <http://eprints.ncrm.ac.uk/3721/>
36. A 2015 recent guide to impact evaluation made the same point. The paper by Elliot Stern, commissioned by the Big Lottery Fund, Bond, Comic Relief and the Department for International Development, said that one of the problems faced by those approaching impact evaluations is that: "it is often presented as a technical or methodological question only accessible to experts or researchers. To some extent this is true but the main arguments, logics and choice-points are more accessible."
37. Department for International Development (2014) 'How to Note: Assessing the Strength of Evidence.' See: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/291982/HTN-strength-evidence-march2014.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf)
38. For more on the importance of anecdote, see Greenhalgh, T. (2014) 'Stories or numbers OR stories and numbers.' See: <http://www.alliance4usefulevidence.org/stories-or-numbers-or-stories-and-numbers/>

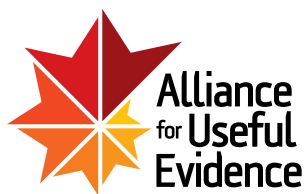


39. For a fundamental critique see Frideres, J. S. (1992) 'Participatory research: An illusionary perspective.' in Frideres, J.S. (ed.) *A world of communities: Participatory research perspectives*. North York, Ontario: Captus University Publications. Frideres has, of course, been heavily criticised by the participatory research community.
40. This is adopted from Table 2, page 18 of Stern, E. (2015) 'Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.'
41. White, H. and Phillips, D. (2012), 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework.' See: [http://www.3ieimpact.org/media/filer\\_public/2012/06/29/working\\_paper\\_15.pdf](http://www.3ieimpact.org/media/filer_public/2012/06/29/working_paper_15.pdf)
42. <http://www.nesta.org.uk/publications/better-public-services-through-experimental-government>
43. Blair, P.S., Fleming, P.J., Smith, I.J., Platt, M.W., Young, J., Nadin, P. et al. (1999) Babies sleeping with parents: case-control study of factors influencing the risk of the sudden infant death syndrome. *British Medical Journal*. 319(7223). 1457-1462.
44. Medical Research Council (2013) 'Developing and evaluating complex interventions: new guidance.' See: <http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>
45. See: <http://www.alliance4usefulevidence.org/publication/spark-guide-to-testing-social-policy-innovation/>
46. See page 5 of Stern, E. (2015) 'Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.'
47. See page 21, *ibid*.
48. Carter, R. (2012) 'Helpdesk Research Report: Theory-based evaluation approach.' Birmingham: Governance and Social Development Resource Centre, University of Birmingham. See: <http://www.gsdr.org/docs/open/hdq872.pdf>
49. Westhorp, G. (2014) 'Realist Evaluation: An Introduction.' London: Methods Lab, Overseas Development Institute. See: <http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf>
50. Mayne, J. (2012) Contribution Analysis: Coming of Age?. *Evaluation*. 18(3). 270-280.
51. For network analysis and process tracing see: Befani, B. and Mayne, J. (2014) Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation. *IDS Bulletin*. 45(6). 17-36. See: <http://onlinelibrary.wiley.com/doi/10.1111/1759-5436.12110/abstract>
52. Baptist, C. and Befani, B. (2015) 'Coffey How To Guide: Qualitative Comparative Analysis – A Rigorous Qualitative Method for Assessing Impact.' See: <http://www.coffey.com/assets/Ingenuity/Qualitative-Comparative-Analysis-June-2015.pdf>
53. See the Scottish Government Social Researchers' Method Guide on Contribution Analysis: <http://www.gov.scot/resource/doc/175356/0116687.pdf>
54. For example the old guide by the late, great evaluation guru Carol Weiss: Weiss, C. (1998) *Evaluation: methods for studying programs and policies*. New York: Prentice Hall.
55. <http://www.thinknpc.org/blog/getting-started-with-cost-benefit-analysis/>
56. Cabinet Office (2007) 'A guide to Social Return on Investment.' See: [https://www.bond.org.uk/data/files/Cabinet\\_office\\_A\\_guide\\_to\\_Social\\_Return\\_on\\_Investment.pdf](https://www.bond.org.uk/data/files/Cabinet_office_A_guide_to_Social_Return_on_Investment.pdf)
57. HM Treasury (2011) 'The Green Book: Appraisal and Evaluation in Central Government.' London: HM Treasury.
58. <https://www.nao.org.uk/report/choosing-the-right-fabric-3/>
59. <https://www.gov.uk/government/publications/supporting-public-service-transformation-cost-benefit-analysis-guidance-for-local-partnerships>
60. Petticrew, M. and Roberts, H. (2003) Evidence, hierarchies and typologies: horses for courses. *Journal of Epidemiology and Community Health*. 57. 527-529.
61. Ioannidis, J.A. (2005) Contraindications and Initially Stronger Effects in Highly Cited Clinical Research. *The Journal of the American Medical Association*. 294(2). 218-228.
62. Cited in *The Economist* (2013) 'Unreliable research; trouble in the lab.' 409(8858). 24.
63. Easterbrook, P.J., Gopalan, R., Berlin, J.A. and Matthews, D.R. (1991) Publication bias in clinical research. *The Lancet*. 337(8746). 867-872.
64. Social sciences suffer from severe publication bias; the survey found that 'null results' rarely see the light of the day.
65. Wolford, B. (2014) 'Big Pharma Plays Hide-the-Ball With Data.' See: <http://europe.newsweek.com/medical-science-has-data-problem-284066?rm=eu>
66. Sense about Science (2006) 'I don't know what to believe: Making sense of science stories.' See: <http://www.senseaboutscience.org/resources.php/16/i-dont-know-what-to-believe>
67. See page 10 of Government Office for Science (2010) 'The Government Chief Scientific Adviser's Guidelines on the Use of Scientific and Engineering Advice in Policy Making.' See: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/293037/10-669-gcsa-guidelines-scientific-engineering-advice-policy-making.pdf)
68. See page 14 of [http://www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised\\_april\\_2015-web.pdf](http://www.evaluationsupportscotland.org.uk/media/uploads/resources/ess-evidenceforsuccess-revised_april_2015-web.pdf)
69. National Audit Office (2013) 'Evaluation in Government.' See: <http://www.nao.org.uk/report/evaluation-government/#>
70. See for instance, see this guide as a supplement to the Magenta Guide: HM Treasury, DECC and DEFRA (2012) 'Quality in policy impact evaluation; understanding the effects of policy from other influences.' The guidance shows how 'higher quality research designs can help meet the challenge of attributing measured outcomes to the policy in question (as opposed to other influences), whereas lower quality designs reduce confidence in whether it was the policy that achieved those outcomes' (page 5).
71. For a list and short description of these scales and hierarchies, see Annex 1 of Nutley, S., Powell, A. and Davies, H. (2013) 'What Counts as Good Evidence?.' London: Alliance for Useful Evidence.
72. For a model of different approaches on quality that includes four dimensions: (1) methodological quality; (2) quality in reporting; (3) appropriateness; (4) relevance to policy and practise see: Boaz, A. and Ashby, D. (2003) 'Fit for purpose? Assessing research quality for evidence based policy and practice.' ESRC UK Centre for Evidence Based Policy and Practice Working Paper 11.
73. See the Research Councils UK and Universities 'UK concordat to support research integrity', signed in 2012 by many key research funders: <http://www.universitiesuk.ac.uk/highereducation/Pages/Theconcordattosupportresearchintegrity.aspx#.Vd2X5fVhBc>
74. Callaway, E. (2011) Report finds massive fraud at Dutch universities. *Nature*. 479(15). See: <http://www.nature.com/news/2011/111101/full/479015a.html>
75. Boaz, A. and Ashby, D. (2003) 'Fit for purpose? Assessing research quality for evidence based policy and practice.' ESRC UK Centre for Evidence Based Policy and Practice Working Paper 11.
76. HM Treasury, DECC and DEFRA (2012) 'Quality in policy impact evaluation; understanding the effects of policy from other influences.'
77. HM Treasury (2011) 'The Magenta Book: Guidance for Evaluation.' London: HM Treasury.
78. HM Treasury (2011) 'The Green Book: Appraisal and Evaluation in Central Government.' London: HM Treasury.
79. Brooks, G., Hall, J. and Torgerson, C.J. (2006) 'A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling.' London: Department for Education and Skills.
80. Henderson, M. (2013) 'The Geek Manifesto: Why science matters.' London: Corgi. Page 17.
81. HM Treasury (2011) 'The Magenta Book: Guidance for Evaluation.' London: HM Treasury.
82. See Morgan, S.L. (ed.) (2013) 'Handbook of Causal Analysis for Social Research.' Berlin: Springer. But it has to be confronted head-on: <http://www.tylervigen.com/spurious-correlations>
83. See page 4 of Stern, E. (2015) 'Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations In the Voluntary and Community Sector.'

84. Qualitative research explores and tries to understand people's beliefs, experiences, attitudes, behaviour and interactions. It generates non-numerical data. Well-known qualitative research methods include in-depth interviews, focus groups, documentary analysis and participant observation, for example.
85. Although over a decade old, the Cabinet Office and defunct Office for the Chief Social Researcher have a guide on the quality of qualitative research. Regardless of where you stand in the paradigm wars, there's enough consensus that quality matters – and that we can't avoid making judgements about what is good or not. See: <http://www.cebma.org/wp-content/uploads/Spencer-Quality-in-qualitative-evaluation.pdf>
86. For an overview – and critique – of hierarchies of evidence, see: <http://www.alliance4usefulevidence.org/assets/What-Counts-as-Good-Evidence-WEB.pdf>
87. Public Health England (2014) 'Identifying what works for local physical inactivity interventions.'
88. [https://www.nesta.org.uk/sites/default/files/nesta\\_standards\\_of\\_evidence\\_csaif.pdf](https://www.nesta.org.uk/sites/default/files/nesta_standards_of_evidence_csaif.pdf)
89. The Nesta and Project Oracle standards were strongly inspired by the independent Social Research Unit at Dartington, who work with many local authorities and children's service providers.
90. Ní Ógáin, E. (2015) 'Impact measurement in impact investing learning from practice.' London: Nesta Impact Investments. See: [http://www.nesta.org.uk/sites/default/files/impact\\_measurement\\_in\\_impact\\_investing.pdf](http://www.nesta.org.uk/sites/default/files/impact_measurement_in_impact_investing.pdf)
91. Ibid
92. Ruiz, F. and Breckon, J. (2014) 'The NICE way; lessons for social policy and practice from the National Institute for Health and Care Excellence.' London: Alliance for Useful Evidence/NICE.
93. Rawlins, M. (2014) Evidence, values and decision-making. 'International Journal of Technology Assessment in Health Care'. 30(2). 233-8.
94. See for example page 28 of Nutley, S. *et al.* (2013) 'What Counts as Good Evidence?'. London: Alliance for Useful Evidence.
95. To see the difference between systematic reviews and meta-analysis, see pages 64-5 of 'The Magenta Book.' For a quick guide on doing reviews, aimed at policymakers see Alliance for useful Evidence (2013) 'Learning from Research: Systematic Reviews for Informing Policy Decisions.': <http://www.alliance4usefulevidence.org/assets/Alliance-FUE-reviews-booklet-3.pdf>
96. For a critique of applying research syntheses into policy, see Pawson, R. (2001) 'Evidence Based Policy: In Search of a Method.' ESRC UK Centre for Evidence Based Policy and Practice: Working Paper 3.
97. See page 9 of <http://www.alliance4usefulevidence.org/assets/Alliance-final-report-08141.pdf>
98. HM Treasury (2011) 'The Magenta Book: Guidance for Evaluation.' London: HM Treasury.
99. Thomas, J., Newman, M., Oliver, S. (2013) Rapid evidence assessments of research to inform social policy: taking stock and moving forward. 'Evidence & Policy: A Journal of Research, Debate and Practice.' 9(1). 5-27.
100. <http://www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/dfe-evidence-check-forum/>
101. Begley, C.G. and Ellis, L.M. (2012) Drug development: Raise standards for preclinical cancer research. 'Nature.' 483. 531–533.
102. <http://www.alliance4usefulevidence.org/replication/>
103. McCormick, J. (2013) 'Evidence Exchange: Learning from social policy from across the UK.' London: Carnegie UK Trust and Joseph Rowntree Foundation.
104. <http://www.nesta.org.uk/blog/why-what-works-centres-are-working>
105. [https://www.nesta.org.uk/sites/default/files/nesta\\_a4ue\\_and\\_nice\\_paper.pdf](https://www.nesta.org.uk/sites/default/files/nesta_a4ue_and_nice_paper.pdf)
106. <http://www.alliance4usefulevidence.org/a-brief-guide-to-uk-social-policy-databases/>
107. <http://www.scotland.gov.uk/Resource/0038/00385880.pdf>
108. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/243718/evidence-reduce-reoffending.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/243718/evidence-reduce-reoffending.pdf)
109. Although this guide is for UK social policy, it's also worth mentioning 3ie short guides in international development. See: <http://www.3ieimpact.org/en/evidence/policy-briefs/3ie-briefs/>
110. <http://www.alliance4usefulevidence.org/rapid-evidence-assessments-a-bright-idea-or-a-false-dawn/>
111. <http://www.alliance4usefulevidence.org/the-creation-of-a-new-service-to-unlock-research-expertise-and-you-hold-the-key/>
112. E.g. European Commission (2010) 'Communicating research for evidence-based policymaking – a practical guide for researchers in socio-economic sciences and humanities.' See: <http://www.spia-europa.de/pdf/guide-communicating-research.pdf>
113. Medical Research Council (2013) 'Developing and evaluating complex interventions: new guidance.' See: <http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>
114. Beynon, P. *et al.* (2012) 'What Difference does a Policy Brief Make?.' International Initiative for Impact Evaluation (3ie), Institute of Development Studies, Norwegian Agency for Development Cooperation. See: [http://www.3ieimpact.org/media/filer\\_public/2012/08/22/fullreport\\_what\\_difference\\_does\\_a\\_policy\\_brief\\_make\\_2pdf\\_-\\_adobe\\_acrobat\\_pro.pdf](http://www.3ieimpact.org/media/filer_public/2012/08/22/fullreport_what_difference_does_a_policy_brief_make_2pdf_-_adobe_acrobat_pro.pdf)
115. [http://www.3ieimpact.org/media/filer\\_public/2012/08/24/3ie\\_policy\\_brief\\_experiment\\_summary.pdf](http://www.3ieimpact.org/media/filer_public/2012/08/24/3ie_policy_brief_experiment_summary.pdf)
116. Behavioural Insights Team (2014) 'EAST: Four Simple Ways to Apply Behavioural Insights.' See: [http://www.behaviouralinsights.co.uk/wp-content/uploads/2015/07/BIT-Publication-EAST\\_FA\\_WEB.pdf](http://www.behaviouralinsights.co.uk/wp-content/uploads/2015/07/BIT-Publication-EAST_FA_WEB.pdf)
117. Research quoted in Royal Mail (undated) 'Door to Door; Big Return. The seven secrets of highly effective leaflet design.' See: [http://www.royalmail.com/sites/default/files/docs/pdf/seven\\_secrets\\_from\\_rm\\_v1.09.pdf](http://www.royalmail.com/sites/default/files/docs/pdf/seven_secrets_from_rm_v1.09.pdf)
118. See Rousseau, D.M. (ed.) (2012) Oxford Handbook of Evidence Based Management. Oxford: Oxford University Press.
119. Allen, G. (2011) 'Early Intervention: The Next Steps An Independent Report to Her Majesty's Government.' See: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/284086/early-intervention-next-steps2.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/284086/early-intervention-next-steps2.pdf)
120. <http://www.theguardian.com/society/2014/may/06/policymakers-neuroscience-justify-early-intervention-agenda-bruce-perry>
121. McCabe, D.P. and Castel, A.D. (2007) Seeing is believing: The effect of brain images on judgments of scientific reasoning. 'Cognition.' 107 (1). 343-352. See: <http://www.imed.jussieu.fr/en/enseignement/Dossier%20articles/article4.pdf>
122. This is part of the Building Capacity to Use Research Evidence (BCURE) programme of work, funded by the UK Department for International Development.
123. <http://www.alliance4usefulevidence.org/the-poor-state-of-evaluation-in-government/>







58 Victoria Embankment  
London EC4Y 0DS

+44 (0)20 7438 2500

[www.alliance4usefulevidence.org](http://www.alliance4usefulevidence.org)  
[www.nesta.org.uk](http://www.nesta.org.uk)

Mae Nesta yn elusen gofrestredig yng Nghymru a Lloegr, rhif cwmni 7706036 a rhif elusen 1144091. Mae'n gofrestredig yn yr Alban, rhif SCO42833. Swyddfa Gofrestredig: 58 Victoria Embankment, London, EC4Y 0DS.

Nesta is a registered charity in England and Wales with company number 7706036 and charity number 1144091. Registered as a charity in Scotland number SCO42833. Registered office: 58 Victoria Embankment, London, EC4Y 0DS.

