Essentiality Rate Inflation and Random Variability in SEP Counts with Sampling and Essentiality Checking for Top-Down FRAND Royalty Rate Setting

Fair, reasonable and non-discriminatory (FRAND) royalty rates for licensing standard-essential patents (SEPs) are increasingly derived "top-down" by dividing a notional aggregate royalty percentage (e.g., 10% of a smartphone's selling price for 4G LTE) among patent owners based on the proportions of patents they own that are deemed to be standard essential.[1] Other rate-setting methods include use of comparable licenses and measuring value derived from SEPs. However, as stated by Justice Birss in *Unwired Planet*: "In assessing a FRAND rate counting patents is inevitable."[2] He used top-down methodology as a cross check. The FRAND rate Decision in *TCL v. Ericsson*, that was unanimously and entirely vacated on appeal, also relied on a top-down valuation.[3]

Exhibit 1: Ball colour, not patent essentiality, can be identified with 100% accuracy when sampling

Essentiality checking is also proposed by the European Commission and others to improve transparency for prospective licensees.[4] But transparency is only legitimate if what is being revealed and counted is reasonably accurate and does not mislead. Otherwise, it will likely do more harm than good.

If patent counting with essentiality checking is going to be used in FRAND-rate determinations, then it is vital we understand its dynamics, failings, how to properly interpret its results, and how to design and size essentiality determination studies that are fit for purpose. This research article contributes to that quest.

---

[1] These percentages in mobile phone licensing are typically applied to the wholesale selling prices of finished goods products.

[2] Approved Judgment in Unwired Planet versus Huawei, 5th April 2017 at 806 (11). https://www.judiciary.uk/wp-content/uploads/2017/04/unwired-planet-v-huawei-20170405.pdf

[3] http://www.ip.finance/2018/04/unreasonably-low-royalties-in-top-down.html and https://www.reuters.com/legal/transactional/ericsson-tcl-settle-long-running-smartphone-patent-disputes-2021-07-19/

[4] https://ec.europa.eu/growth/content/transparency-sep-licensing-how-clarify-possible-exposure-upfront_en

Figuring out which patents are true SEPs and then counting them is fundamental to top-down analysis. With hundreds of thousands of patents declared possibly standard essential,[5] and the insuperably huge task in checking them all for essentiality, the European Commission and others propose that assessing only samples of patents declared essential to a standard would suffice.

Institutionalising use of patent counting—with or without sampling—is an under-researched and impetuous leap of faith. Different assessors come up with wildly different patent counts. Comparison of two separate assessors determining essentiality on the same sample of patents indicates that assessors tend to inflate true essentiality rates. These over-estimates result from statistical bias with numbers of false positives (i.e., truly not essential patents being found to be essential) exceeding false negatives when true essentiality rates are rather less than 50%.

I also conclude from my simulation modelling and analysis that:

1. The lower the true essentiality rate and the lower the rate of agreement among different assessors, the larger the differences will be between the essentiality rates determined by assessors and the true essentiality rate. For example:

    a. If true essentiality rates are 30% and two different assessors agree with each other on 75% of their determinations, they will tend to estimate 36% essentiality rates and be accurate in 85% their determinations.

    b. If true essentiality rates are only around 10% (e.g., for 4G LTE or 5G), as some experts plausibly argue, and if two assessors agree with each other on 84% of their determinations, they will tend to estimate 17% essentiality rates and be accurate in 91% of their determinations. That means there will be nearly as many false positives as correct determinations of essentiality.

2. Therefore, if true essentiality rates are at the lower end of expectations, for example, at around 10% or less, it is imperative assessors are highly accurate in their determinations, otherwise false positives will swamp their correct determinations and make their overall results meaningless.

Sampling reduces the precision of SEP counts with increased variability, which is exacerbated by erroneous essentiality determinations. Sampling theory and simple simulations using results of patent-counting studies already undertaken—including several with sample sizes below a few hundred—reveal unacceptably large ranges in expected essentiality rate determinations (i.e. the percentage of declared-essential patents that are deemed to be essential)[6] at what various study authors regard as the "well accepted bound" of the 95% confidence level.[7] This variability is particularly large where patent essentiality rates are at low levels, such as at around 10%. For example, 10% ± 1.5% is actually ± 15% variability as a proportion of that 10% figure. The quantitative analysis I undertook for this article measures the extent of diminutions, which should be properly and fully considered before sample sizes are set, and before the short cut of sampling is blindly adopted at all.

[5] Declaring one's patents that are possibly standard essential is a requirement for participation in organisations such as 3GPP in the setting of standards such as 4G LTE and 5G.
[6] This is also called the essentiality ratio.
[7] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3752684 at page 4 and https://media.crai.com/wp-content/uploads/2020/09/16163609/Transparency_predictability_efficiency.pdf at page 38

Top-down methodology seemingly enables precise assessments of FRAND royalties, but this is an illusion due to various inconsistencies and inaccuracies in patent selection, sampling and essentiality assessment. Key questions are how much precision is adequate and how can that be obtained? The optimal balance between extensive and costly patent-essentiality assessments, that can take days of work per patent, versus reducing the number of assessments by sampling should be decided with due regard to accuracy and confidence levels required, and based on empirical assessments.

While there are no set bounds for the acceptably accurate range in determinations, I have considered a reasonable proportionate accuracy requirement for essentiality rate determination to be <± 15% (i.e., a 30% range for the determined essentiality rate as a proportion of the true essentiality rate) at the 95% confidence interval level. With my opinion that true essentiality rates are more like 10% than 30% or 40%, I conclude from my analysis that samples including thousands of patents are required in top-down FRAND-royalty rate setting. For example, if the essentiality rate is only 10%, a sample size approaching 3,000 declared-essential patents per standard, at the very least, would be required.

Valuation methods, alleged royalty stacking and pie sharing

While there are several ways to determine licensors' FRAND royalty charges for SEPs— including from comparable licenses, and by measuring the incremental product and service value bestowed by the patented technologies—top-down valuation is increasingly considered in licensing negotiations and litigation. However, top-down has never been accepted anywhere in litigation as a standalone valuation method.

Top-down methodologies including an aggregate royalty-rate cap and SEP counts are used as a cross-check for other rate-setting methods and are proffered to fix the purported problem that cumulative royalties would otherwise be or become excessive due to alleged "royalty stacking". For example, the European Commission's 2017 Communication *Setting out the EU approach to Standard Essential Patents* states that:

> *"to avoid royalty stacking, in defining a FRAND value, an individual SEP cannot be considered in isolation. Parties need to take into account a reasonable aggregate rate for the standard"*.[8]

Repudiating stacking allegations, my analysis elsewhere shows that cumulative royalties paid are in low single digit percentages and have declined in recent years despite there being numerous patent owners, an exponential increase in SEP declarations and the introduction of new standard such as 5G over the last couple of years.[9]

Raw declared-essential patent counts are rightly considered inapplicable to top-down FRAND-rate setting due to significant numbers of declared-essential patents not being truly essential. Therefore, SEPs counted must only be those found to be standard essential in accurate and reliable independent assessments.

Despite patent counting being antithetical to patent law, top-down analysis is appealing to the courts and is increasingly used by them because it is a practical solution that makes the maths simple and provides the illusion of being equitable. Top-down methods are generally on the simplistic and erroneous basis that all SEPs are of equal value. On the contrary, evidence indicates

---

[8] COM(2017) 712 final, at Page 7. https://ec.europa.eu/docsroom/documents/26583

[9] https://www.rcrwireless.com/20210903/analyst-angle/modest-sep-royalties-on-smartphones-have-declined-and-licensing-is-stabilizing

there are massive disparities in value among patents[10]. Some patents have much stronger links to technical contributions and specifications for foundational new technologies than other declared-essential patents.

Patent counting also typically ignores the major question of patent validity, with no independent assessment of that or any adjustment of patent counts for patents that would be judged invalid in court. An invalid patent cannot be truly standard essential, and so true essentiality rates must be lower than estimated to the extent to which a court would find patents invalid. Might patent counts be more overstated by declared-essential patents simply being invalid, than by them being found not essential while validity is ignored?[11] Another major empirical assessment would be required to answer that question and measure the extent to which invalidity also reduces the numbers of patents that are truly essential.

In my previous articles, I have critically analysed the selection of aggregate royalty cap figures in detail, along with my critical analysis of top-down FRAND determinations in general.[12] I am not repeating all that here because my descriptions and analysis on top-down are context for my focus on using patent sampling in top-down FRAND-rate setting, and resulting effects on assessment accuracy. This article is not intended to be a "soup to nuts" critique on top-down methodologies.

No particular top-down valuation method has been ordained, but divvying-up a notional aggregate royalty rate based on companies' share of patents deemed to be standard essential is increasingly common in litigation. In the Innovatio case of 2013, Judge Holderman used the term to describe the method in which he valued the patented technologies as a proportion of the profits that could be derived in manufacture of chips implementing those technologies.[13]

More commonly and prevalently since then, top-down has meant allocating a maximum aggregate royalty rate among licensors based on SEP counts. This is a simple and relatively easy way of deriving royalty rates. The methodology regards FRAND royalty rates to be directly proportional to the shares of SEPs owned by licensors.[14] For example, if, hypothetically, the maximum aggregate royalty for a standard such as 4G LTE is deemed to be 10% of smartphone selling prices and a licensor is found to own 10% of patents that are truly essential to that standard, then that licensor's maximum royalty charge before cross-licensing offsets would be 10% x 10% = 1% of smartphone selling prices.

In theory, top-down methodology only requires accurate SEP counts, with a lower requirement for accuracy on individual patent assessments. Erroneous essentiality determinations for some patents would not matter if these errors are, on average, offset by opposing determination errors for other patents. In these circumstances, every false positive would be offset by a false negative and the total count of SEPs would be correct. It might seem that this helpful netting-off of determination errors

---

[10] For example, Judge Holderman's Innovatio case Opinion in 2013 cited expert witness testimony "finding that the top 10% of all electronics patents account for 84% of the value in all electronics patents". https://www.essentialpatentblog.com/wp-content/uploads/sites/64/2013/10/2013.10.03-975_Public-Version-of-Memorandum-Opinion-and-Order.pdf

[11] Research by Mark Lemley and Timothy Simcoe in 2019 indicates lower essentiality rates than validity rates, but that was for litigated SEPs. That might not be the case for declared SEPs in general. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3128420

[12] http://www.ip.finance/2018/04/unreasonably-low-royalties-in-top-down.html

[13] https://www.essentialpatentblog.com/2013/10/public-version-of-judge-holdermans-rand-determination-in-innovatio-wifi-sep-litigation/

[14] Most top-down royalty allocations, including by most patent pools and in recent litigation including Unwired Planet and TCL v. Ericsson (unanimously and entirely reversed on appeal) use this approach, at least as a cross check with other valuation methods.

should ensure a good level of accuracy; but any such presumption must be empirically tested in case false negatives and false positives are in significantly different proportions, as appears to be the case, particularly where essentiality rates are much lower than 50%. In other words: all well and good if the total of patents counted is correct. But that is a very big if. I suspect counts of patents—declared and then found to be essential—and corresponding essentiality rates estimated in practice tend to be way too high, and various other experts agree, as I will show in this article.

Serendipitous SEP declarations?

The goal of standard setting organisations (SSOs) such 3GPP and its partners including ETSI is to produce standards based on the best technical solutions which are widely accessible and will not be blocked by standard-essential patents not being available for license under FRAND terms and conditions. The possibility of being able to use SEP counts to apportion royalty rates stems from the obligation participants to these and other SSOs have in patent or intellectual property rights (IPR) policies to declare patents that potentially are or might become standard essential. Counting SEPs in top-down royalty rate analysis is using a by-product of this process. For example, the purpose of 3GPP partner ETSI's IPR policy is to "reduce the risk that the investment in the preparation . . . of [standards] could be wasted as a result of an [essential] IPR . . . being unavailable"[15] and that, "IPR holders . . . should be adequately and fairly rewarded for the use of their IPRs."[16] By implementing a FRAND licensing policy, SSOs seek to prevent SEP holders from refusing to license their SEPs on reasonable terms—thereby denying implementers access to the technology. At the same time, SSO IPR policies aim to ensure that SEP holders will be fairly and adequately compensated for their contributions to the standard. There is nothing there that anticipated using the count of SEPs in the setting of royalty rates.

Patent claims and standard specifications often change after the declaration. Consequently, some legitimately declared patents end up being non-essential. That does not make them "over-declarations." [17] However, that does not preclude other patents being over-declared.

With increasing use of SEP counts to determine royalty rates in valuation methods, there are major incentives to "game the system" and so it is inevitable that patent owners will increasingly declare patents that never end up being standard essential. For example, net payers of royalties are motivated to inflate the total base of SEPs by amassing large portfolios for themselves and thus diluting SEP shares for others. Human and organisational behaviour here, including over-declarations, exemplifies Goodhart's Law: '"When a measure becomes a target, it ceases to be a good measure." In other words, when we set one specific goal, people will tend to optimize for that objective regardless of the consequences.'[18] I have heard of examples of declared-essential patent technology appearing to read on a standard technical specification, where the patent or application priority date is later than that for publication of the first technical specification version employing that technology. So that patent should not be found valid; but it does boost an individual's and a company's declaration figures and might catch out an assessor doing essentiality analysis. This example highlights why ignoring validity may further prejudice patent counts.

Patent counting is also commonly used when determining royalty rates from "comparable licenses". SEP counts are used to derive the patent "portfolio strength ratio" in an algebraic formula often employed to "unpack" cross-licenses and derive "one-way" licensing rates from these.[19]

[15] https://www.etsi.org/images/files/IPR/etsi-ipr-policy.pdf.
[16] https://www.etsi.org/images/files/IPR/etsi-ipr-policy.pdf.
[17] This term has pejoratively come to mean declaration of any patents that are not truly essential.
[18] https://towardsdatascience.com/unintended-consequences-and-goodharts-law-68d60a94705c
[19] https://www.iam-media.com/frand/frand-royalty-rates-in-sep-licensing-comparable-licence-agreements

Transparency on essentiality and royalty costs

Essentiality checking is also proposed by the European Commission and others to improve transparency for prospective licensees.[20] But transparency is only legitimate if what is being revealed and counted is reasonably accurate and does not mislead. Otherwise it will likely do more harm than good.

Transparency is worse than useless if the information provided is biased, distorted, prejudiced or otherwise inaccurate. Such information can provide a spurious basis for decision-making and give highly-misleading impressions. One example of this kind of shortcoming was in 2007, when the Next Generation Mobile Network (NGMN) Alliance, an industry group led by mobile operators and including major 4G equipment vendors, established a confidential process for the ex-ante disclosure and aggregation of expected licensing fees for a number of upcoming 4G standards including LTE.

The stated objective was "to increase transparency on royalty rates" for 4G LTE to members by adding up the maximum licensing terms of would-be licensors.[21] The aggregate figures calculated and presented to members in tens of percent were far higher than the single digit figures, which would ever actually be paid, because the analysis did not reflect the results of negotiation, for example, with identification of weak portfolios and cross-licensing.[22] These shortcomings were inevitable and easily foreseen by licensing experts. The process concluded in 2009. I have suspected for many years that the process was purposely concocted to produce such distorted results; so as to put pressure on those who are dependent on licensing income to accept lower patent fees.[23]

My concern about patent counting is that it also tends to inflate figures, with large and increasing numbers of SEPs crowding out and diluting the perceived value of truly essential and valuable patents. Institutionalising the process, with the European Commission proposing to set up a centralised process with some kind of ministry of patent counting, will entrench and reinforce this undesirable effect. Mechanisms and outcomes might also be subjected to political pressures.

Reasonable minds may differ

With differences in opinion on essentiality and with differences in how conservative or opportunistic patent owners are with declarations, as well as differences in the nature of patent portfolios, estimated essentiality rates may vary considerably among assessors. Declared SEPs are commonly subjected to independent essentiality checks. This is a requirement for patent pools, where there is a legal obligation that only standard-essential patents are licensed. Similarly, patent counts used in top-down analysis outside of pools, for example, in published studies, commonly employ essentiality checks to adjust counts for declared-essential patents that are not deemed essential.

The difficulties in determining which or how many patents are standard-essential stems from the fact that assessing essentiality for any patent is laborious, full of uncertainties and includes subjective decision making. The only way for essentiality to be definitively determined for any

---

The use of patent counting to unpack cross-licenses is not considered further in this article that focuses on top-down analysis. Nevertheless, this topic is also worthy of further consideration because FRAND rates derived from unpacked cross-licenses are also highly sensitive to inaccuracies in patent counts.

[20] https://ec.europa.eu/growth/content/transparency-sep-licensing-how-clarify-possible-exposure-upfront_en

[21] Quote was from an NGMN brochure, the link to which is no longer active.

[22] http://www.ip.finance/2021/09/modest-sep-royalties-on-smartphones.html

[23] http://www.ip.finance/2015/08/cumulative-mobile-sep-royalty-payments.html

patent is through court judgments including appeals. But even then, as the courts have noted in recent cases:

> "Based on my assessment of both experts, I am sure the disagreement represents cases in which reasonable people can differ."[24]
>
> "Given the somewhat subjective nature of these determinations, 'disagreements' is probably a more accurate label than 'error.'"[25]

So, it is fantasy that patent essentiality can ever be established with absolute certainty, even for small numbers of patents.[26] Nevertheless, we will stick with the notion that there is a truth about whether a patent is actually essential. The test for that is what the applicable courts would rule if it ever came to that. However, it is only for rather less than 1% of patents that the courts ever do that. For most patents, essentiality has not been and never will be determined by the courts. That makes it impractical ever to measure the accuracy of any patent assessor against this rather nebulous norm. Instead, next best in seeking to find out how accurate essentiality assessments are, is to compare the essentiality determinations of different assessors—including counts and shares of deemed-essential patents, and where possible, comparing such assessments on a patent-by-patent basis.

I have already shown elsewhere that the patent counting is full of inaccuracies, inconsistences and disagreements among different assessors. In an article I published in 2017, I found wide disparities among studies in found-essential shares of companies' patents that read on the 4G LTE standard.[27] While published studies rank companies by share of found-essential SEPs owned, most of these do not reveal their assessments on a patent-by-patent basis.

As I have already explained, patent counting for top-down FRAND royalty rate determinations, it does not, in theory, matter how many individual errors there on what is and what is not essential if false negatives cancel out false positives to yield the correct essentiality rate.

Nevertheless, despite any mitigation from netting-off such errors, there are remarkably wide variations in results among different studies. This is due to reasonably transparent and mechanical differences in which patents are selected for assessment, as well as the subjective determinations on which of these are essential.

Exhibit 2: A few among many wide variations in shares of found-essential LTE patents various among patent-counting studies

	Lowest Estimate	Highest Estimate	Disparity
Huawei	2.9%	23%	8x
LG	0.6%	17%	17x
Nokia	2.3%	54%	23x

Source: WiseHarbor 2017[28]

---

[24] Unwired Planet Judgment at 335. https://www.judiciary.uk/wp-content/uploads/2017/04/unwired-planet-v-huawei-20170405.pdf

[25] TCL v. Ericsson at FN 16

[26] That is also the case regarding product infringement and validity.

[27] https://www.wiseharbor.com/wp-content/uploads/2017/05/Patent-Counting-article-for-IP-Finance-12-May-2017.pdf

[28] https://www.wiseharbor.com/wp-content/uploads/2017/05/Patent-Counting-article-for-IP-Finance-12-May-2017.pdf

There have also been significant disparities in patent-by-patent essentiality determinations undertaken in litigation. Two different experts who were both retained on behalf of TCL in TCL v. Ericsson disagreed about essentiality for many patents, as shown in Exhibit 3. The large extent of this is remarkable, given that both experts were batting on the same side and claimed to be coordinating their assessment activities.[29]

Exhibit 3: Disagreement between Dr Ding's (Concur IP) and Dr Kakaes's essentiality determinations in TCL v. Ericsson

	2G	3G	4G	TOTAL
Disagree	2	3	11	16
Agree	5	13	25	43
Total	7	16	36	59
Disagreement rate (%)	29%	19%	31%	27%
Agreement rate (%)	71%	81%	69%	73%

Henceforth, I will call these agreement rates "Consistency Rates". By doing that I am using the same terminology as does the European Commission's 2020 Pilot Study for Essentiality Assessment of Standard Essential Patents in its comparison of different essentiality assessors.[30] A Consistency Rate of 100% means agreement on every essentiality determination, a rate of 0% means that assessors disagree on their determinations for every single patent.

A rate of 50% means there is no relationship or correlation between the assessors' determinations. This is as if at least one of the two assessors was determining essentiality based on the outcome of a coin toss (e.g., heads for essential and tails for not essential)[31]. There are four permutations of equal probability: heads and heads, heads and tails, tails and heads, tails and tails. That means that the two assessors agree with each other on 50% of essentiality assessments, purely by chance.

Which patents are selected for assessment, who deems which are essential and how?

They key issue at hand is in comparing essentiality assessments and patent counts derived from all patents or samples of them in those categories. However, major differences also stem from how the relevant patents for assessment get to be selected in the first place, including:

- Are patent applications as well as issued patents included?
- What range of dates are applicable for declarations, patent priorities, lapsed applications, issued and expired patents?
- How are patent transfers dealt with, if at all?
- Which patent nationalities and languages are considered?
- How is the term patent family interpreted and how many family members might be counted?
- Are patents applicable to user equipment, network equipment or systems?
- Are optional and obsolete as well as mandatory parts of the standards included?

[29] https://www.essentialpatentblog.com/wp-content/uploads/sites/64/2018/01/2017.12.21-1802-Court-Memo-of-Facts-and-Law-PUBLIC-CORRECTED.pdf
[30] European Commission, Joint Research Center, "Pilot Study for Essentiality Assessment of Standard Essential Patents," *EUR 30111 EN*. https://publications.jrc.ec.europa.eu/repository/handle/JRC119894
[31] http://www.ip.finance/2018/04/unreasonably-low-royalties-in-top-down.html

- Which standard releases and technical specifications, including revised, withdrawn and replaced versions are applicable?
- How are patents that are declared essential to more than one generation of the standards dealt with?

While much of the above may be somewhat transparent, studies and their assessors commonly use different selection criteria to each other, which will affect SEP counts even before essentiality determinations are made on selected patents.

There is no gold standard and there are no stipulations about exactly what is applicable for assessment and inclusion in court-judged essential or found essential patents.

While I accept that definitions and use of terms varies among authorities and commentators, for clarity in this article, I use the following terms to describe SEPs and their essentiality determinations most specifically:

1. Declared essential—before any check on essentiality independent of the declarer.
2. Truly essential—cannot be known for all declared-essential patents, or any of them with certainty, but what a court would judge (even hypothetically) is tantamount to true or actual essentiality.
3. Found essential—what has been determined by a third-party assessor.

Differences between the patent counts for 2. and 3. are thus in exactly what is assessed for essentiality, as well as in the subjective assessments on those.

Some essentiality studies use more nuanced terminology and categorisation than the binary alternatives of *essential* and *not essential*. Terms include patents *actually, really, truly, in fact, verified, probably* and *appear essential,* and whether essential patents are *seminal or highly novel.* In *TCL v. Ericsson,* TCL's expert Dr. Kakaes rated patents with a score of 1, 2 or 3 with respect to essentiality, depending on claim construction.

As stated previously, the major matter of validity is typically ignored, even in what is regarded as truly essential or what would be judged essential in a court of law.

Thorough assessments on essentiality can take many hours and include use of claim charts and review of patent prosecution histories. Preparation of claim charts can take several days per patent. Some studies make rapid assessments of essentiality without these references in as little as less than one hour per patent family. The European Commission's Pilot Study concluded:

> *"The best results (i.e. most consistent with patent pool outcomes) are achieved by individuals working at a patent office as patent examiners, who are provided with a claim chart, and perform a regular essentiality assessment (as opposed to a novelty-based assessment). They achieve a consistency rate of 84% with the pool outcomes, and report spending 5.9 hours on average per assessment."*[32]

Again, and notably, no assessment of validity was attempted.

[32] European Commission, Joint Research Center, "Pilot Study for Essentiality Assessment of Standard Essential Patents," *EUR 30111 EN*, at page 81. https://publications.jrc.ec.europa.eu/repository/handle/JRC119894

Top and bottom

The two fundamental issues in top-down royalty-rate setting are selecting the aggregate royalty rate and determining the subject licensor's share of patents that are deemed to be truly essential to a particular standard. While I have much to say about the first issue, and have already done so along with criticising top-down FRAND rate determinations in general,[33] as already mentioned, this article focuses the shortcomings of SEP sampling in setting top-down royalty rates.

A patent owners' share of SEPs is simply defined as the number of SEPs it owns divided by the number of SEPs in the "universe" or "pool" of all SEPs reading on a particular standard, such as 4G/LTE or 5G.

Also according to the 2020 report on the European Commission's Pilot Study for Essentiality Assessment of Standard Essential Patents:[34]

> "'*numerator data'*, is information on the actual SEPs portfolio by of a specific patent owner for a specific standard. This data type is relevant for all stakeholders and **requires a high degree of confidence**."
>
> *"'denominator data'*, is information on actual SEPs owned by *all* relevant patent owners for a specific standard. In combination with numerator data, it can indicate the size (extent) of the SEPs portfolio owned by a specific patent owner in relation to all SEPs for the standard in question. Being able to do so is crucial for one of the licensing principles expressed by the European Commission, which states that, in defining a FRAND value, an individual SEP cannot be considered in isolation, and one needs to take into account a reasonable aggregate rate for the standard. This data type is relevant for all stakeholders, **and only requires a moderate level of confidence, which could also be satisfied by sampled data."**
>
> (Emphasis added, citation omitted.)

Seemingly easy with regard to the denominator, and with only one number required? We do not need to know exactly which patents are and are not essential, all we need is a patent count for the particular standard or standards. But this is not that easy or trivial, even to do it one-off for one standard, let alone to deal with the complexities with multiple, and ever-evolving standards with exponential growth in patents and essentiality declarations over the years.

I disagree that one can jump to the conclusion that a lower level of confidence is acceptable in the denominator and that sampling can be summarily assumed to be satisfactory for that. While one patent-essentiality determination error in the denominator will have a less significant effect on a patent owner's estimated share of SEPs than one single such error in the numerator, an error *rate* (e.g., of 1 in 10 determinations) will be just as significant in the numerator as in the denominator. Careful consideration of results from quantitative sensitivity analysis is needed to measure and interpret diminished confidence levels on resulting FRAND royalty rate determinations.

While commercial studies tend to assess all SEPs, or a sample of them, in the same way, it is common in FRAND-rate litigation to employ a different and more thorough assessment methodology for the subject company's SEPs in the numerator, to those in the entire universe of SEPs in the denominator. For example, in TCL v. Ericsson, for the numerator, of 311 patent (families) declared essential, consideration was only included a subset of 112 of them for which Ericsson had produced

[33] http://www.ip.finance/2018/04/unreasonably-low-royalties-in-top-down.html
[34] https://publications.jrc.ec.europa.eu/repository/handle/JRC119894 at page 117.

claim charts. Only 70 of these were not challenged for essentiality by TCL. For the denominator, in a completely separate exercise, TCL's expert Dr Ding in conjunction with a firm called Concur IP estimated that 1,796 out of 5,024 declarations were essential. Another of TCL's expert reduced this figure to 1,673 to account for regional differences and then the Court further reduced numbers for "over-declaration" to 1,481 (i.e. by 11%).

Judge Selna indicated the Court's desire for a "workable size of the relevant universe" for the denominator, but, as indicated above, derived rates are just as arithmetically sensitive to proportionate changes in the denominator as they are to the same proportionate changes in the numerator.

The "baby-splitting" Justice Birss understandably resorted to in setting the denominator reveals severe inconsistency and inadequacy in prospective figures provided to him by the parties that differed by a factor of four. In the same timeframe and also with focus on LTE user equipment (UE) patents in Unwired Planet, Justice Birss computed top-down rates based on his rough and readily estimated "pool" of only 800 patents in the denominator.[35] Ericsson's share of SEPs would have been virtually twice Judge Selna's assessment in TCL v. Ericsson using this figure.

Correspondingly, the above difference in the size of denominators between the two cases' top-down calculations largely accounts for an approximately factor of two disparity in top-down royalty rates derived by the different courts in these two cases.

The difference between Justice Birss' and Judge Selna's denominators also accounts for the difference between Unwired Planet's satisfaction with Briss' FRAND determinations and Ericsson's profound disappointment with Judge Selna's FRAND determinations, that were subsequently vacated on appeal.

A factor of two discrepancy between those two denominator judgments is unacceptably large. However, I recognise that such measurements cannot be made with the high precision, for example, of using a tape measure to measure a table with nominal length of two metres, to an accuracy of ±1 mm (i.e. ± 0.005% of the table's length).

For this research article, I am working on the basis that a fit-for-purpose patent-essentiality and FRAND-rate determination system requires that total variability and inaccuracies should be no more than around 30% (i.e. ± 15% of a central figure) at the 95% confidence interval level. I would not disagree if some were to argue that "tolerance"[36] level is not good enough. I am merely putting a stake in the ground so I have a figure to work with, to focus thought on what might be minimum acceptable levels of accuracy and what is required to achieve them in design and sizing of patent counting studies.

Mixing and mismatching

It is generally bad practice to calculate proportions with numerators and denominators derived from different methodologies. This is an apples to oranges comparison. I agree with the European

---

[35] Unwired Planet Decision, paragraph 377. "The significant overstatement in the HPA is the number produced for the total pool of Relevant SEPs. The number for 4G handsets is 1812 and is much too high. The corresponding number in the Revised MNPA is 355 but that number is much too low if it is to represent all Relevant SEPs. I think both values are out by about a factor of two. Half of 1812 is 906 while twice 355 is 710. Splitting the difference takes one to 800. Standing back, about 800 is fair and in my judgment an appropriate figure for the pool of 4G/LTE patents. Applying that as the denominator in a fraction to determine the share…" https://www.judiciary.uk/wp-content/uploads/2017/04/unwired-planet-v-huawei-20170405.pdf

[36] I use the term as do by engineers: https://fractory.com/engineering-tolerances/

Commission's 2020 report on its Pilot Study for Essentiality Assessment of Standard Essential Patents on this issue.[37] If the same patent selection criteria, essentiality determination method and assessors are used to derive both numerator group and denominator group figures, across-the-board errors and statistical biases such as major differences in which patents are assessed, misplaced "benefits of the doubt" on essentiality may tend to cancel out because they might inflate or deflate the numerator and denominator to the same extent. If in selecting declared-essential patents for which to assess, and then making essentiality determinations is more rigorous for the smaller numbers of them in the numerator than the denominator (e.g., giving more benefit of the doubt in the latter), this will tend to dilute resulting SEP share percentages downwards. In other words, it is unfair and prejudicial to apply a higher standard of assessment to one group of patents than to the other.

Distortions will occur if assessors are allowed or encouraged to communicate with patent owners or colleagues at all in order to ask for clarification, to consult additional information sources such as the patent prosecution histories —about one group of patents, but not about the other. Even if such inquiries lead to increased accuracy in one group of assessments, if there is no corresponding improvement in the other group of assessments, the accuracy in determining the essentiality rates will reduce and the results will be prejudicial. This is one case where two wrongs might tend to cancel each other out and keep the overall assessment closer to being right.

While claim charts can help establish essentiality, availability of these of these for assessing patents in the denominator will always be deficient. Claim charts really need to be prepared voluntarily by those organisations who declare essentiality and can most effectively advocate for the patents to be deemed essential.[38] The lower the essentiality rate (i.e., with essentiality assessors seeking to separate the wheat from the chaff), the more important this advocacy becomes. To make persuasive claim charts, one needs to be convinced that claims read on the standard and find readings in the patent and claims that support this view. Assistance from the inventor who also knows the standard is the fastest, easiest and most effective way of making a claim chart demonstrating essentiality.

Preparing claim charts is a time consuming task that can take several days per patent. Ericsson testified that it spent 50 hours per patent preparing claim charts in TCL v. Ericsson. Eleven hours per patent were spent preparing claim charts and determining patent essentiality in Dr David Cooper's 2019 study on 4G LTE SEPs entitled Evaluating Standard Essential Patents in Mobile Cellular.[39] Claim charts will never be created for most declared-essential patents. It also seems unlikely that claim charts could timely be obtained, even in small quantities, from all patent declarers, as would be required with random sampling of patents for the denominator in patent counting studies.

The promise of AI

Automated methods including Artificial Intelligence (AI) will be increasingly useful tools to assist in selection of patents to assess, and in essentiality determinations, but as with manual methods, efficacy and reliability of such techniques must also be verified before using, let alone relying on AI for FRAND-rate determinations. Analysing this topic is outside the scope of this article. However, I

---

[37] Id at page 88, "we suggest the assessment of the chosen sample [for the denominator] is of precisely the same methodology and carried out in the same rigor as that for numerator data to prevent systematic biases." And on page 105, "when sampling is used, the actual resulting assessments must be using same methodology and same rigor for both numerator and denominator data, otherwise they are incomparable."

[38] The EC's Pilot Study states that "The availability of claim charts (made available by the patent owner) as input for an assessment procedure is an important aspect of designing a system that combines high quality with high efficiency."

[39] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3470197

note that the EC Pilot Study Report at page 13 indicates that automated approaches will not be able to replace human efforts for full essentiality assessments in the short or medium term.

The lure and shortcomings of sampling

One significant shortcoming of patent-counting studies is that assessments tend to be only cursory because there are too many patents to assess. Published studies and studies conducted by or relied upon by expert witnesses in FRAND-rate litigation have typically spent less than an hour per patent in finding whether the overall universe of patents are standard essential. Assessments on rather smaller numbers of patents for prospective inclusion in patent pools, in patent-infringement litigation and in assessing the parties' patents in FRAND-rate litigation typically take orders of magnitude more time per patent, and commonly include reviews of detailed claim charts and patent prosecution file histories.

While assessing only a random sample of the total number of patents could substantially reduce the number of patents to be assessed and enable the sampled patents to be assessed in much greater depth, there has, so far, been no full and proper review of this alternative, including quantification of additional inaccuracies introduced by sampling and assessing only that subset of patents. In other words, the trade-offs have not been properly weighed or counterbalanced.

If patents' relative values, as well as patent counts, were also, properly, taken into account, the statistical variances and inaccuracies through sampling become even greater, and dramatically so. The difference between including or not including a most highly valued patent in a sample would be enormous. This is not a reason disregard patents' relative values, it is a reason to ensure that high-high value patents are identified and are recognized when setting royalty rates. For example, Cisco has paid CSIRO around one dollar per device for licensing one SEP to the IEEE 802.11 WiFi standard.[40] Average royalties per patent paid to license the WiFi standard have always been orders of magnitude lower. If the values of a small number of SEPs massively exceed all others in a standard, the only representative sample is one that includes all patents.

Empirical analysis on essentiality studies using random sampling

Study results can mislead about confidence intervals due to how readers might misinterpret figures and due to outright egregious error.

For example, in analysing 5G SEPs in 2021, Dr David Cooper et al:

> *"estimate that 746 of the 9,327 non-Ericsson patent families are likely essential to the Release 15 5G NR smartphone, and with a confidence level of 95%, this provides a likely essential rate of 8.0% ± 3.7% for non-Ericsson patent families."*[41]

At a glance, ± 3.7% might seem like a small margin of error due to sampling, but it indicates a range of essentiality rate from 4.3% to 11.7%. That is a factor of 2.8 difference between those two percentages and ± 47% of the 8.0% central figure. Any supposedly FRAND royalty rates derived from these metrics would have similar variability. While FRAND royalty rates cannot be determined exactly, this level of imprecision is woefully low.

[40] http://www.cafc.uscourts.gov/sites/default/files/opinions-orders/15-1066.Opinion.12-1-2015.1.PDF

[41] Survey of Mobile Cellular 5G Essentiality Rate, by David Cooper, Johanna Dwyer and Alexander Haimovic, February 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3771397

Charles River Associates' 2016 *Report for the European Commission* on *Transparency, Predictability, and Efficiency of SSO-Based Standardization and SEP Licensing* states at page 61:[42]

> *"Still, there is one additional issue to consider in a random testing environment: we have to ask how intensive the sampling of portfolios needs to be to provide us with a reliable estimated. There is a rigorous statistical answer to that question. Either a patent is essential, or it is not. Denote the proportion of essential patents in a portfolio of K patents by α. How many of the K patents do we need to test so that the proportion of patents found essential, defined as β is likely to be close enough to β? Suppose for example that we test 30 patents and find that 30% of them are essential. Using the "normal approximation" approach to the distribution of a binomial sample mean, we would get that there is a 95% chance that the actual proportion of truly essential patents in the whole portfolio is between 27% and 33%. This is quite a good precision so that the method would not expose patent-holders to any considerable risk of error."*

CRA got its maths homework wrong here: it massively overstated the precision and confidence level that can be obtained from its proposed sample size. Using the same binomial theory equation I used to verify Dr David Cooper et al's 95% confidence interval figures, CRA's chosen sample size and essentiality rate would yield a range in essentiality rate of 30% ±16.4%.[43] That is a range of 13.6% to 46.4% for the essentiality rate at the 95% confidence interval level. Increasing the sample size by a factor of 10 and reducing the confidence interval to 68% (i.e., within one standard deviation rather than 1.96 standard deviations of the 30% mean) yields the 27% to 33% range claimed by CRA.

Furthermore, simplistic application of simple binomial theory understates the inaccuracies that occur in the real world in essentiality analysis of sampled patents.[44] What is fundamental, but overlooked when study authors trot out theoretical confidence interval figures like these is the implicit but faulty assumption that sampled patents can be identified as essential or not essential with 100% certainty, as is the case in sampling other items. In contrast, imagine randomly selecting a sample of balls from a bag containing only red and white balls. So long as one has decent eyesight and the lighting is adequate, one would be able to identify the colour of sampled balls with 100% certainty. In contrast, when patents are sampled, determination of essentiality is subject to significant uncertainty that determinations will be incorrect, or different from what a court or other competent assessor would make. While there are also those same uncertainties when assessing all declared SEPs rather than merely a sample of them, the smaller the sample, the greater statistical impact on essentiality rate range for any given confidence interval level. The range between the lower essentiality rate and the higher essentiality rate bounds for the 95% confidence level, or for any other confidence level, is wider than predicted by binomial sampling theory alone.

It is an empirical question as to how much wider that range will be. This can be numerically modelled in simulations. I have done this using various figures from recent research studies to measure the diminution due to real-world sampling. But before doing all that, it is useful to consider how estimation accuracy is affected in sampling, only by sample size and by the level of the essentiality rate.

---

[42] https://media.crai.com/wp-content/uploads/2020/09/16163609/Transparency_predictability_efficiency.pdf

[43] $\sigma=(E(1-E)/N)^{0.5}$: where σ= standard deviation of error due to sampling, E=essentiality rate and N= sample size https://onlinestatbook.com/2/sampling_distributions/samp_dist_p.html

[44] The shortcoming is more basic and significant than the symmetrical normal probability density function in binomial theory sampling being an imperfect fit to situations where the distribution is *skewed* non-normal (i.e., asymmetric), for example, due to E being close to 0%.

Simplistic and understated inaccuracies with binomial theory

Exhibit 4 shows the confidence intervals predicted purely on the basis of the binomial theory – as if patent essentiality is determined with 100% accuracy in agreement with what a court would judge if it assessed every sampled patent for essentiality. As discussed above, that is far from being a valid assumption – particularly for studies that spend little time per patent, for example, only 20 minutes by Concur IP for Dr Ding in TCL v. Ericsson—in assessing relatively large numbers of patents for essentiality (e.g., 2,600 in that case). I consider the effects of errors and inaccuracies in essentiality determination in the next section. The 95% confidence interval (C.I.) is within 1.96 standard deviations (i.e., 1.96 x σ) of the mean (i.e., the estimated essentiality rate).

Exhibit 4: Various recent essentiality studies using random patent sampling

	Cooper 2019 LTE[45]	**Cooper et al 2021 5G**	**CRA 2016 hypothetical**	**CRA 2016 adjusted: 10x sample**	**Concur IP TCL v. Ericsson**	**EC Pilot 2020 experiment**
Sample size N	200	200	30	300	2,600	205
Essentiality rate E	12%	8.0%	30%	30%	37.3%	30%
Standard deviation σ additive	2.30%	1.92%	8.37%	2.65%	0.95%	3.20%
1.96 x σ additive (95% confidence)	4.5%	3.76%	16.4%	5.2%	1.9%	6.3%
Lower bound, 95% C.I.	7.5%	4.2%	13.6%	24.8%	35.4%	23.7%
Higher bound, 95% C.I.	16.5%	11.8%	46.4%	35.2%	39.2%	36.3%
Higher/Lower 95% C.I.	**2.20**	**2.77**	**3.41**	**1.42**	**1.10**	**1.53**
1.96 x σ/p proportionate (95% confidence)	**37.5%**	**47.0%**	**54.7%**	**17.3%**	**5.0%**	**20.9%**
Lower bound, 68% C.I.	9.7%	6.1%	21.6%	27.4%	36.4%	26.8%
Higher bound, 68% C.I.	14.3%	9.9%	38.4%	32.6%	38.2%	33.2%
Higher/Lower 68% C.I.	**1.47**	**1.63**	**1.77**	**1.19**	**1.05**	**1.24**
σ/p proportionate (68% confidence)	**19.1%**	**24.0%**	**27.9%**	**8.8%**	**2.5%**	**10.7%**

Source: WiseHarbor

Nevertheless, the relatively large sample size of 2,600 used by Concur IP yields a relatively small variability (i.e., confidence interval range at the 95% confidence level) of ± 5.0% <u>of</u> the 37.3% essentiality rate, purely based on sampling theory.

Others might argue that a ± 20.9% variability at the 95% confidence level on the EC Pilot Study's assumed essentiality rate of 30% makes a sample size of as small as 205 adequate.[46] However, the variability as a proportion of the much lower essentiality rates of 12% and 8% in the two Cooper studies approximately doubles to unquestionably unacceptably high levels of ± 37.5% and ±47.0% respectively, despite all three studies' sample sizes being nearly the same.

Therefore, the approximate level of the essentiality rate anticipated, as well as the sample size are both very important factors in designing an essentiality assessment using patent sampling. There is universal agreement that the overall essentiality rates for patents declared essential to cellular

[45] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3470197
[46] Id at page 80: "For illustration, assume that 30 percent of a sample of declared SEPs are actually essential".

standards are substantially less than 100%, and there appears to be consensus that these rates are no more than 50% overall (even if some companies are found by some assessors to have substantially higher essentiality rates), but as the sources in Exhibit 4 indicate, there is wide range of opinion on the approximate level of that percentage, from approaching 40% down to below 10%.[47]

Pluralities of inconsistencies and inaccuracies in SEP counts

Considering only inaccuracies introduced by sampling for the denominator is deficient. Inaccuracies are introduced in the denominator and in the numerator from patent sampling and with imprecise or erroneous essentiality determinations and disagreements between these and what a court would judge to be standard essential.

Even if all of a subject company's declared-SEPs are assessed for essentiality (i.e., no sampling), all the essentiality assessments for the numerator and for the denominator are subject to essentiality determination errors and patent count inaccuracies.

An experiment in the EC Pilot Study measured the extent to which two different expert assessors come to the same conclusion on whether declared patents are truly essential.

> *"The most consistent results are achieved by individuals who work at a patent office as patent examiners and are provided with a claim chart. They achieve a consistency rate of 84% (while spending considerably less time than the pool assessors). In our experiments, assessors that work as senior engineer in academia score below that (75% consistency rate, without input claim charts)."*[48]

The Pilot Study experiment's approach ought to result in the closest possible agreement between any two assessors. Comparing the assessments of two different experts with very similar professional backgrounds on the same sample of patents seems more likely to result in determination agreements than if assessors had different professional backgrounds and independently decided which patents to select for assessment.

These essentiality assessments took approximately 7 hours each. It seems unlikely that consistency would increase, even versus a court assessment that was "better", by being more thorough and taking a lot longer (i.e., akin to what occurs in patent infringement litigation) while the other assessor still only took 7 hours. While the assessors achieving the 84% consistency rates were provided with claim charts, these would not be available (not even created, let alone provided) for many patents selected randomly in assessment of the denominator of the "universe", or "pool", of all patents declared essential to the standard.

While consistency rates of 75% and 84% might seem fairly high, it should be recognised that assessors would have an expected 50% consistency rate even if they made their essentiality judgments based on the toss of a coin, as I have already explained.

---

[47] Essentiality rates for individual patent owners and for all of them overall in other studies are also very varied. For example, Fairfield Resources (2010) estimated essentiality rates among LTE SEP declarers ranging from 18% to 67% for those declaring more than five patents, with an average of 50% across all declarers. Studies on LTE SEPs in three consecutive years by Cyber Creative Institute (2011, 2012 and 2013) found several declarers to have essentiality rates in excess of 80%, and overall essentiality rates between 50% and 60%. A report analysing 5G SEP declarations to 2019 by Amplified and GreyB found essentiality rates ranging from 19% to 34% for six named companies, an essentiality rate of 28% for "Others" and an overall essentiality rate of 26%.

[48] Id at page 116.

With the essentiality rate as low as 10%, such seemingly high consistency rates could be even more misleading. For example, two assessors could both come up with 10% essentiality rates by agreeing on all of which 80% of patents are not essential, while also being in total disagreement on essentiality for the all the other 20% of patents.

This Case A example is illustrated in Exhibit 5 along with two other extreme cases in comparing SEP determinations on a sample of 50 patents. A binary 1 denotes that a patent is determined to be essential and a binary 0 denotes that it is determined to be not essential. Case B is where two different determinations have the same 10% essentiality rate by agreeing on all determinations. Case C is an example where one of the assessors determines a 10% essentiality rate and the other is in agreement on 90% of determinations, but with no agreement that any patents are essential. The order of the patents is irrelevant in any of these cases because there is no interdependence between patents. I have placed the essential determinations towards the top of the list to make it easier to see how and where there are agreements and disagreements on essentiality between assessors.

It is mathematically impossible for two different assessors to both determine an essentiality rate of 10% or less without having a consistency rate of at least 80%. Imagine, as in the extreme Case A below, there is agreement on non-essentiality for 80% of patents and disagreement on essentiality for all of the other 20% of patents. Any more agreements would increase the Consistency Rate. With total disagreement about which patents are essential, any additional disagreement (i.e., lowering the Consistency Rate) would increase the number of patents found essential by one assessor, which would increase the essentiality rate it finds.

Exhibit 5: Extreme cases in essentiality comparisons

	Case A			Case B			Case C		
	Essentiality Determination			Essentiality Determination			Essentiality Determination		
Patent #	**X**	**Y**	**X & Y Agree?**	**X**	**Y1**	**X & Y1 Agree?**	**X**	**Y2**	**X & Y2 Agree?**
1	1	0	0	1	1	1	1	0	0
2	1	0	0	1	1	1	1	0	0
3	1	0	0	1	1	1	1	0	0
4	1	0	0	1	1	1	1	0	0
5	1	0	0	1	1	1	1	0	0
6	0	1	0	0	0	1	0	0	1
7	0	1	0	0	0	1	0	0	1
8	0	1	0	0	0	1	0	0	1
9	0	1	0	0	0	1	0	0	1
10	0	1	0	0	0	1	0	0	1
11	0	0	1	0	0	1	0	0	1
12	0	0	1	0	0	1	0	0	1
13	0	0	1	0	0	1	0	0	1
14	0	0	1	0	0	1	0	0	1
15	0	0	1	0	0	1	0	0	1
16	0	0	1	0	0	1	0	0	1
17	0	0	1	0	0	1	0	0	1
18	0	0	1	0	0	1	0	0	1
19	0	0	1	0	0	1	0	0	1
20	0	0	1	0	0	1	0	0	1
21	0	0	1	0	0	1	0	0	1
22	0	0	1	0	0	1	0	0	1
23	0	0	1	0	0	1	0	0	1
24	0	0	1	0	0	1	0	0	1
25	0	0	1	0	0	1	0	0	1
26	0	0	1	0	0	1	0	0	1
27	0	0	1	0	0	1	0	0	1
28	0	0	1	0	0	1	0	0	1
29	0	0	1	0	0	1	0	0	1
30	0	0	1	0	0	1	0	0	1
31	0	0	1	0	0	1	0	0	1
32	0	0	1	0	0	1	0	0	1
33	0	0	1	0	0	1	0	0	1
34	0	0	1	0	0	1	0	0	1
35	0	0	1	0	0	1	0	0	1
36	0	0	1	0	0	1	0	0	1
37	0	0	1	0	0	1	0	0	1
38	0	0	1	0	0	1	0	0	1
39	0	0	1	0	0	1	0	0	1
40	0	0	1	0	0	1	0	0	1
41	0	0	1	0	0	1	0	0	1
42	0	0	1	0	0	1	0	0	1
43	0	0	1	0	0	1	0	0	1
44	0	0	1	0	0	1	0	0	1
45	0	0	1	0	0	1	0	0	1
46	0	0	1	0	0	1	0	0	1
47	0	0	1	0	0	1	0	0	1
48	0	0	1	0	0	1	0	0	1
49	0	0	1	0	0	1	0	0	1
50	0	0	1	0	0	1	0	0	1
Total	**5**	**5**	**40**	**5**	**5**	**50**	**5**	**0**	**45**
Essentiality rate	**10%**	**10%**		**10%**	**10%**		**10%**	**0%**	
Consistency Rate			**80%**			**100%**			**90%**

How large do samples need to be?

The European Commission's 2020 Pilot Study for Essentiality Assessment of Standard Essential Patents analysed various issues and advocated use of patent sampling in its 160 page report,[49] but it ducked the question of trying to figure out how large any samples of declared-essential patents for assessment should be:

> *"As it is beyond the scope of this report to calculate the required sample size for a specific context, we adopt a working assumption of a 10% sampling size for the purpose of indicative calculations. Whether that 10% is a proper number will require more investigation."*[50]

Even though the above 10% sampling size figure seems to be little more than a guess with the sample number only implied (e.g., based on total numbers of patents or patents and patent applications declared essential), a percentage is not the pertinent metric. Accuracy in estimating essentiality rates (i.e., the determination range for these at any given confidence level) is dependent on the absolute number rather than the percentage of items sampled from a rather larger universe of declared-essential patents, as I show in my use of the widely-accepted binomial sampling theory equation.

IPlytics and TU Berlin found for the 5G standard alone, there were 95,526 patent and patent application declarations, and 21,571 declared families by the beginning of 2020.[51] Assuming one patent per family is representative of all family members for the purposes of essentiality determination and patent counting in top-down royalty rate analysis,[52] a 10% sample therefore requires more than 2,157 patents.

According to Dr Stec, as cited by David Cooper et al, there are 8,699 patent families declared to 5G with a US or EP member in the 5G census.[53] A 10% sample of those would be 870.

The rest of this article analyses how large samples need to be, based on sampling theory and with various numerical simulations. Those who are not inclined to review this may prefer to skip to the conclusions.

Simulation for sampling and imperfect essentiality determinations

First I simulated the random effects that I have modelled algebraically with the binomial theory above. Secondly, I modelled the probability that any given essentiality determination is either correct or incorrect (i.e., in agreement with what the courts would decide is truly essential or vice versa, respectively). My simulations—run in an Excel spreadsheet workbook—also use a binary representation of essentiality.

---

[49] The report included the caveat on page 105 that "when sampling is used, the actual resulting assessments must be using same methodology and same rigor for both numerator and denominator data, otherwise they are incomparable."

[50] Id at page 96.

[51] *Fact finding study on patents declared to the 5G Standard,* by IPlytics and TU-Berlin, January 2020. https://www.iplytics.com/wp-content/uploads/2020/02/5G-patent-study_TU-Berlin_IPlytics-2020.pdf

[52] This a most optimistic assumption. It is beyond the scope of this article to discuss in detail why reality differs and by how much, however, for example, patent claims may vary among family members, and this impacts essentiality.

[53] *A survey of Mobile Cellular 5G Essentiality Rate,* by David Cooper, Johanna Dwyer and Alexander Haimovich, February 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3771397

From a base of all patents declared essential to a given standard, my model considers four different essentiality rates:

1. The True Essentiality Rate (TE), which cannot be determined exactly because patents (let alone all of them) cannot be correctly identified with certainty as being essential or not essential.
2. The Sample Essentiality Rate (SE) which differs from ET due to variations in sampling, but which also cannot be determined with certainty, as above.
3. The Found Essentiality Rate (E1), which is the estimated value of TE imperfectly determined by a first assessor.
4. The Found Essentiality Rate (E2), which is the estimated value of TE imperfectly (and independently of the first assessor) determined by a second assessor.

My model quantifies relationships between TE and E1, and between TE and E2. It twice simulates agreements and disagreements between an assessor and actuality (i.e., which patents would be judged essential and judged not essential), and it also checks for agreement and disagreement between assessors on an patent-by-patent basis. I model individual imprecise determinations (e.g., disagreements), averages and standard deviations of results. Each assessor's determinations have a similar relationship with actuality, but their determinations are subject to random effects and so they make some different determinations, but with the same averages (i.e. Found Essentiality Rates) and standard deviations in their results.

I use two different measures of consistency. Test results in the real world such as in the EC Pilot Study's *experiment* produce values for E1 and E2 and the rate of agreement, the Consistency Rate C, between the two assessors. In addition, I define the Accuracy Rate A as the extent of agreement between either assessor and actuality. My model uses the same values of C and A for *essential* determinations as for *not essential* determinations. This means that rates of disagreement about what is essential are the same as those for what in not essential.

While the model can accommodate any values of TE, E1, E2, C and A from 0% to 100%, our attention should be focused on values below 50% for TE, E1 and E2, and above 50% for C and A. Most pertinent figures for review are essentiality rates in the range of 8% to 37.5%, as indicated in Exhibit 4, together with Consistency Rates in the range of 75% and 84% and up to the 90% figure "illustration" provided in the EC Pilot Study's experiment and report.[54]

Sampling was modelled patent-by-patent with use of Excel's RAND function [=IF(RAND()<"TE",1,0)]. The RAND function generates a 10-digit random number between 1 and 0. With 1 representing essential and 0 representing not essential, the essentiality rate ES is the total count of 1s generated divided by the total number of patents in the sample.

I used binomial theory predictions to test my simulation model with sample sizes from 200 to 3,000 and averaged results (including of standard deviations) over 1,000 immediate simulations. I found close agreement in mean and standard deviation figures, as can be seen in comparing sampled mean and standard deviation figures from my simulations in the following section with the figures I calculated algebraically there and in Exhibit 4. Repeatedly re-running the all the simulations shows similarly close agreement.

Essentiality assessments tend to inflate essentiality rates

[54] Id at pages 116 and 80.

The values of E1 and E2 tend to be inflated, with statistical bias towards 50%, versus any value of True Essentiality Rate TE below 50% (i.e., in our zone of interest). This is because false positives exceed false negatives with those values of TE.[55] No wonder over-declaration is popular and appears to be something of an "arms race". Even when patent portfolios are checked for essentiality, false positives tend to the boost counts of SEPs found to be essential.

Both E1 and E2 have the same bias so there is symmetry between those values. The asymmetry is between the assessors' biased perceptions of the actuality in TE that are reflected in E1 and E2.

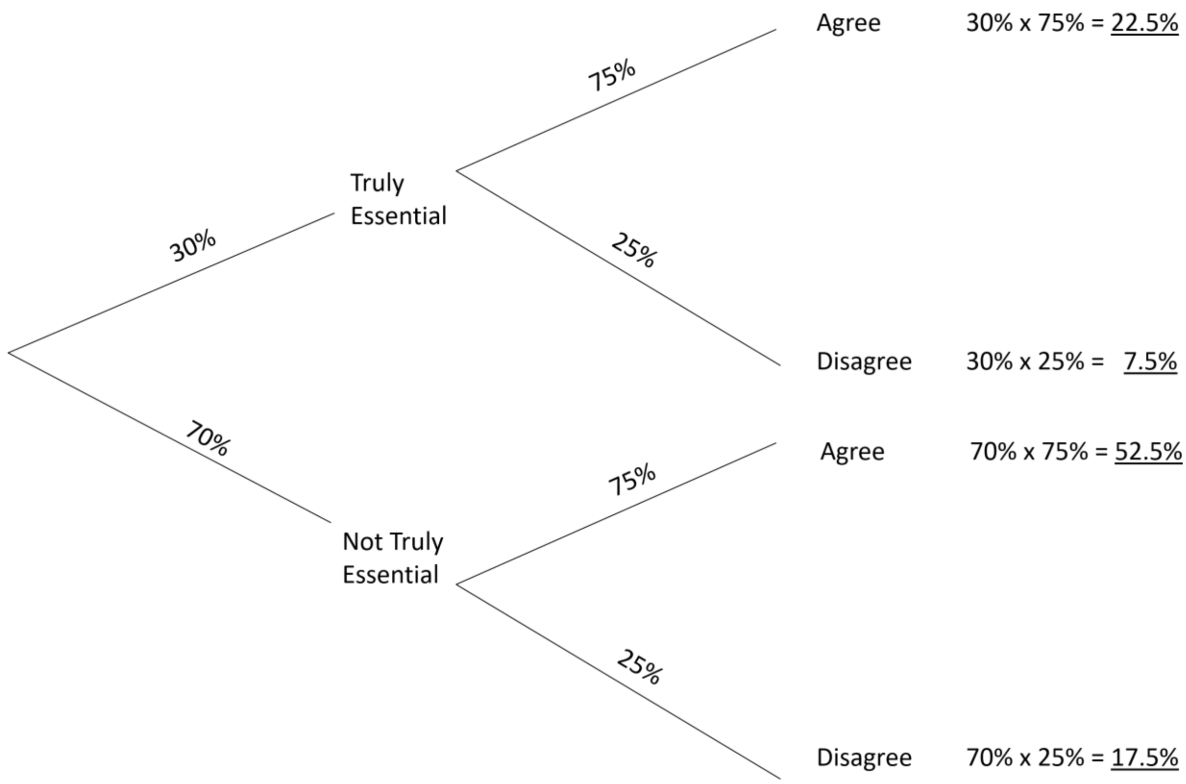
The differences are substantial at an essentiality rate of 30% with the 75% Consistency Rate presented in the EC's Pilot Study. While sampling does not change the mean level of the essentiality rate, determining essentiality on the basis of these two parameters increases the Found Essentiality Rate (E1 or E2) over the True Essentiality Rate by around 10%.

In our range of interest, where essentiality rates are below 50% and Consistency Rates C are above 50%, assessors tend to over-estimate essentiality by a substantial margin. For example, if the probability of a patent being truly essential is 30% (i.e., TE=30%) and the probability of accurate assessment (for both essentiality and non-essentiality) is 75% (i.e., A=75%), the Found Essentiality Rate E1=E2 is: 30%x75%+(1-30%)x(1-75%)=22.5%+17.5%=40%. That is 10% more than than TE=30%.

The probability tree in Exhibit 6 depicts this and the probability of all other possible outcomes underlined with total probability for all outcomes of 22.5% + 7.5% + 52.5% +17.5% = 100%.

[55] The EC Pilot Study report recognizes this at page 80.

Exhibit 6: Probability tree for selecting patents and determining their essentiality



This bias is because the lower the True Essentiality Rate TE, the greater the impact of false positives over false negatives— if, as we assume, the Accuracy Rate A is the same for an assessor's *essential* determinations as on its *not essential* determinations. The difference between the True Essentiality Rate TE or the Sampled Essentiality Rate SE and the Found Essentiality Rates E1 or E2 becomes proportionately larger with lower TE and SE figures.

In my model, an essentiality determination was also made on every sampled patent, also using Excel's RAND function [=IF(RAND()<"A","SAMPLE",BITXOR("SAMPLE",1))], where SAMPLE is the binary value depicting whether a sampled patent is essential or not essential in the previous formula.[56]

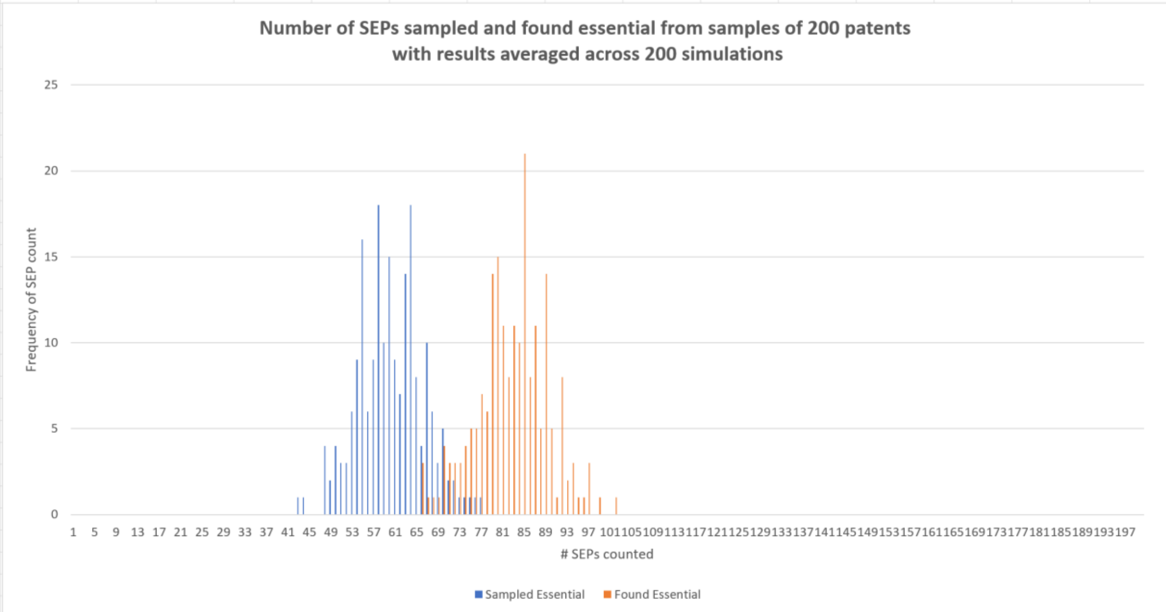
This biasing effect is shown with histograms from simulations including 200 patents and 1,000 patents in Exhibits 7a and 7b. These histograms also show that the biasing becomes more severe at lower essentiality rates.

In my example with TE=10%, I have also increased the Accuracy Rate A because, as I explained previously, it is mathematically impossible for the essentiality rates to align at that percentage or below without the agreement rate (i.e. Accuracy Rate A) being at least 80%. At these rates, false positives massively outnumber false negative. It means there are around as many false positives as correct essentiality determinations.

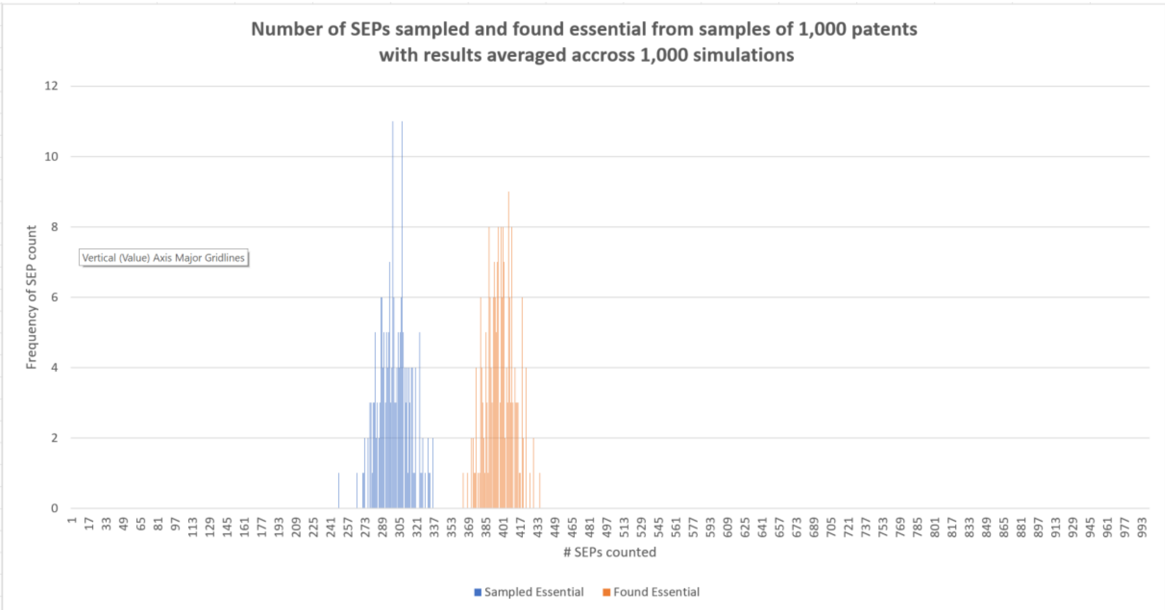
[56] The BITXOR function with one input variable fixed is used to synthesize the logical NOT function because Excel does not have a BITNOT function.

The biasing effect is also illustrated in a tabular form in Exhibit 8 from a very small sample of 50 patents where the True Essentiality Rate TE is 10% and the Accuracy Rate A is 90%. For there to be no bias (i.e., the number of false negatives equal the number of false positives) where the True Essentiality Rate TE is only 10%, the Accuracy Rate A for *not essential* determinations would need to be an improbably nine times higher than that for determinations that patents are found essential by the assessor.[57]

Exhibit 7a: Proportion of found-essential patents biases towards 50%



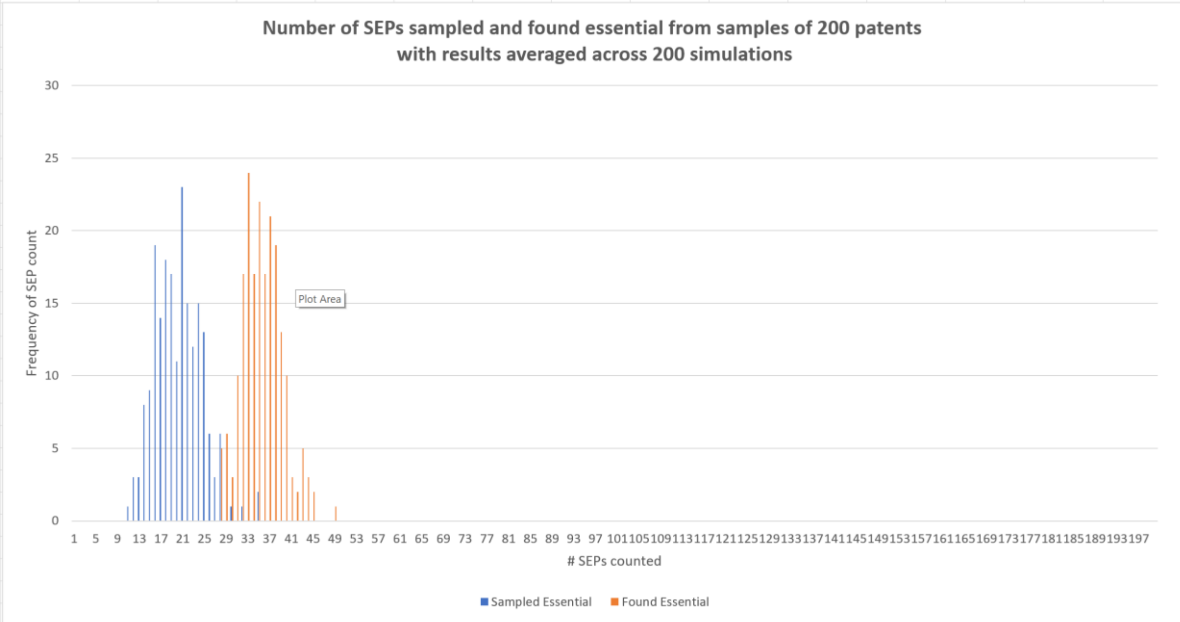
Accuracy Rate	75.0%		True Essentiality Rate	30.0%		Sampled Essentiality Rate	30.1%		Found Essentiality Rate	41.3%



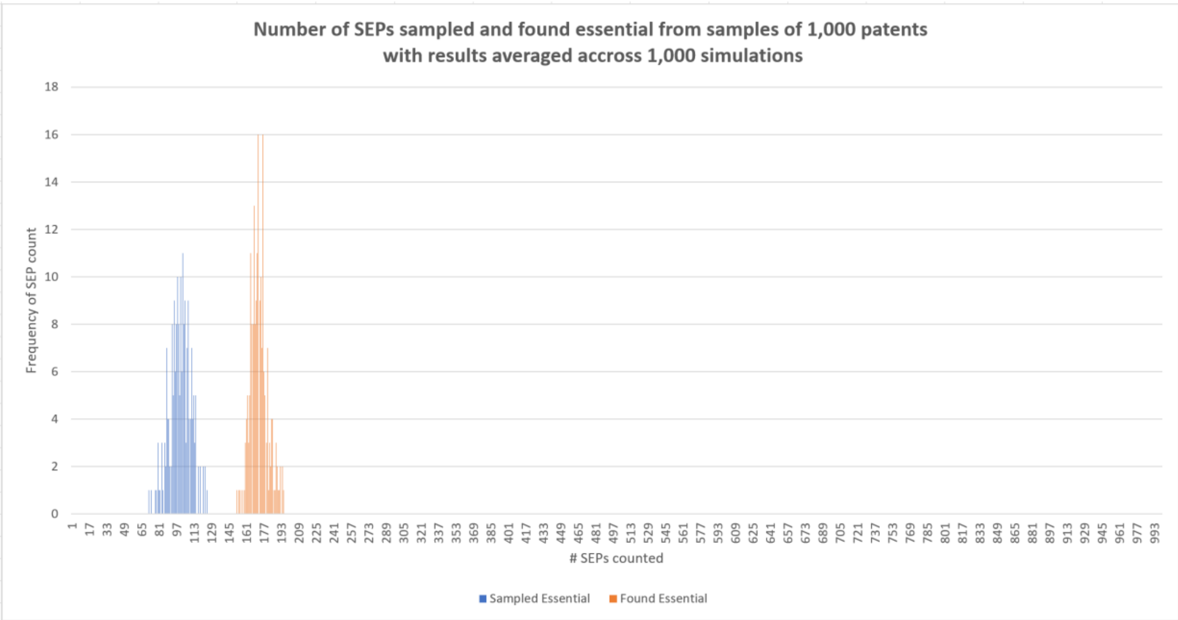
Accuracy Rate	75.0%		True Essentiality Rate	30.0%		Sampled Essentiality Rate	30.1%		Found Essentiality Rate	39.8%

[57] With these dynamics and the use of SEP counts in FRAND determinations there is an unsurprising incentive to "over-declare" patents that are questionably essential in the expectation some will pick up some false-positive essentiality determinations.

Exhibit 7b: Proportion of found-essential patents biases towards 50%



Accuracy Rate	91.0%		True Essentiality Rate	10.0%		Sampled Essentiality Rate	10.1%		Found Essentiality Rate	17.8%



Accuracy Rate	91.0%		True Essentiality Rate	10.0%		Sampled Essentiality Rate	10.0%		Found Essentiality Rate	17.3%

Exhibit 8: Typical results from three small-scale simulations of Found Essential patents

	Case D			Case E			Case F		
	Essentiality Determination			Essentiality Determination			Essentiality Determination		
Patent #	X	Simulation with 90% Accuracy Rate set	X & Simulation Agree?	X	Simulation with 90% Accuracy Rate set	X & Simulation Agree?	X	Simulation with 90% Accuracy Rate set	X & Simulation Agree?
1	1	0	0	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1
4	1	1	1	1	0	0	1	1	1
5	1	1	1	1	1	1	1	1	1
6	0	0	1	0	1	0	0	0	1
7	0	0	1	0	0	1	0	0	1
8	0	0	1	0	0	1	0	0	1
9	0	0	1	0	0	1	0	0	1
10	0	0	1	0	1	0	0	0	1
11	0	1	0	0	0	1	0	0	1
12	0	0	1	0	0	1	0	0	1
13	0	0	1	0	0	1	0	0	1
14	0	0	1	0	0	1	0	0	1
15	0	1	0	0	0	1	0	0	1
16	0	0	1	0	0	1	0	0	1
17	0	0	1	0	0	1	0	1	0
18	0	0	1	0	0	1	0	0	1
19	0	0	1	0	0	1	0	0	1
20	0	0	1	0	0	1	0	0	1
21	0	0	1	0	0	1	0	1	0
22	0	0	1	0	0	1	0	0	1
23	0	0	1	0	0	1	0	1	0
24	0	0	1	0	0	1	0	0	1
25	0	0	1	0	0	1	0	0	1
26	0	0	1	0	0	1	0	0	1
27	0	0	1	0	1	0	0	0	1
28	0	0	1	0	0	1	0	0	1
29	0	0	1	0	0	1	0	0	1
30	0	0	1	0	0	1	0	0	1
31	0	0	1	0	0	1	0	0	1
32	0	0	1	0	0	1	0	0	1
33	0	0	1	0	1	0	0	0	1
34	0	1	0	0	0	1	0	0	1
35	0	0	1	0	0	1	0	0	1
36	0	1	0	0	0	1	0	0	1
37	0	0	1	0	0	1	0	0	1
38	0	0	1	0	0	1	0	0	1
39	0	0	1	0	0	1	0	0	1
40	0	0	1	0	0	1	0	0	1
41	0	0	1	0	0	1	0	0	1
42	0	0	1	0	0	1	0	0	1
43	0	0	1	0	0	1	0	0	1
44	0	0	1	0	0	1	0	0	1
45	0	0	1	0	0	1	0	0	1
46	0	0	1	0	0	1	0	0	1
47	0	0	1	0	0	1	0	0	1
48	0	0	1	0	1	0	0	0	1
49	0	0	1	0	0	1	0	0	1
50	0	0	1	0	0	1	0	0	1
Total	**5**	**8**	**45**	**5**	**9**	**44**	**5**	**8**	**47**
Essentiality Rates	**10%**	**16%**		**10%**	**18%**		**10%**	**16%**	
Accuracy Rates			**90%**			**88%**			**94%**

Imperfect essentiality determinations also increases variability in results

Adding essentiality determinations to sampling also increases the overall standard deviation in results, as how in Exhibits 9 and 10.

Exhibit 9: Compounding effects of sampling and essentiality determination with relatively high essentiality rates

True Essentiality Rate TE	30%
Found Essentiality Rate E1 and E2	36%

Accuracy Rate A	85%
Consistency Rate C	75%

Samples of 200 patents with results averaged over 1,000 simulations

	Sampled Essentiality Rate SE			**Sampled and Found Essentiality Rate E1 or E2**		
	Mean	**σ**	**1.96 x σ**	**Mean**	**σ**	**1.96 x σ**
	29.92%	3.23%	6.32%	36.07%	3.42%	6.70%
Binomial theory prediction		*3.24%*				
Min @ 1.96 x standard deviation			23.6%			29.4%
Max @ 1.96 x standard deviation			36.2%			42.8%
Difference (i.e., range @ 95% C.I. level)			12.6%			13.4%
Difference as percentage of TE			42.2%			44.7%
± percentage of TE (i.e., half of the above)			21.1%			22.3%

Samples of 1,000 patents with results averaged over 1,000 simulations

	Sampled Essentiality Rate SE			**Sampled and Found Essentiality Rate E1 or E2**		
	Mean	**σ**	**1.96 x σ**	**Mean**	**σ**	**1.96 x σ'**
	30.00%	1.47%	2.88%	36.00%	1.56%	3.07%
Binomial theory prediction		*1.45%*				
Min @ 1.96 x standard deviation			27.1%			32.9%
Max @ 1.96 x standard deviation			32.9%			39.1%
Difference (i.e., range @ 95% C.I. level)			5.8%			6.1%
Difference as percentage of TE			19.2%			20.4%
± percentage of TE (i.e., half of the above)			9.6%			10.2%

Variability (i.e., the range of the Found Essentiality Rate at any given Confidence Interval level), as percentage of the True Essentiality Rate, increases as the True Essentiality Rate is reduced.

Exhibit 10: Compounding effects of sampling and essentiality determination with relatively low essentiality rates

True Essentiality Rate TE	10%
Found Essentiality Rate E1 and E2	17%

Accuracy Rate A	91%
Consistency Rate C	84%

Samples of 200 patents with results averaged over 1,000 simulations

	Sampled Essentiality Rate SE			**Sampled and Found Essentiality Rate E1 or E2**		
	Mean	**σ**	**1.96 x σ**	**Mean**	**σ**	**1.96 x σ**
	10.01%	2.07%	4.06%	17.19%	2.82%	5.52%
Binomial theory prediction		*2.12%*				
Min @ 1.96 x standard deviation			5.9%			11.7%
Max @ 1.96 x standard deviation			14.1%			22.7%
Difference (i.e., range @ 95% C.I. level)			8.1%			11.0%
Difference as percentage of TE			81.3%			110.4%
± percentage of TE (i.e., half of the above)			40.6%			55.2%

Samples of 1,000 patents with results averaged over 1,000 simulations

	Sampled Essentiality Rate SE			**Sampled and Found Essentiality Rate E1 or E2**		
	Mean	**σ**	**1.96 x σ**	**Mean**	**σ'**	**1.96 x σ'**
	9.97%	0.95%	1.86%	17.19%	1.17%	2.30%
Binomial theory prediction		*0.95%*				
Min @ 1.96 x standard deviation			8.1%			14.9%
Max @ 1.96 x standard deviation			11.8%			19.5%
Difference (i.e., range @ 95% C.I. level)			3.7%			4.6%
Difference as percentage of TE			37.1%			46.0%
± percentage of TE (i.e., half of the above)			18.6%			23.0%

Samples of 3,000 patents with results averaged over 1,000 simulations

	Sampled SE			**Sampled and Found Essentiality Rate E1 or E2**		
	Mean	**σ**	**1.96 x σ**	**Mean**	**σ**	**1.96 x σ**
	9.99%	0.54%	1.06%	17.19%	0.68%	1.33%
Binomial theory prediction		*0.55%*				
Min @ 1.96 x standard deviation			8.9%			15.9%
Max @ 1.96 x standard deviation			11.1%			18.5%
Difference (i.e., range @ 95% C.I. level)			2.1%			2.7%
Difference as percentage of TE			21.3%			26.7%
± percentage of TE (i.e., half of the above)			10.6%			13.3%

Across the gamut of essentiality rates, assessor accuracy and consistency rates

My model results measure the statistical bias across the entire range of essentiality rates, assessor Accuracy Rates A and Consistency Rates C. Exhibit 11 is a tabulation of Found Essentiality E1 or E2 for all possible combinations of True Essentiality Rate TE and Accuracy Rate A in the range of 0% to 100% for both variables. To guide our focus on the applicable ranges of essentiality and precision, I have shaded-out the three quadrants of the chart that are not pertinent in the real word (other than for mathematical curiosity). I have also highlighted the couple of examples already discussed in yellow.

There is an additional complication. We do not have a direct measure of the Accuracy Rate A. Instead, we have from the EC Pilot experiment and other studies, measures of the extent to which different assessors agree with each other (i.e., the Consistency Rate C).

As explained previously, my simulation model reflects the different determinations of these two assessors. They each have an identical relationship with actuality in their determinations. They are simply represented by two different simulations, between which I can also compare results. Modelled on this basis, their determinations tend to be in closer agreement with what is truly essential than they are with each other's essentiality assessments. That is to be expected: they are assessing what they believe is essential and what they believe is not be essential. They are not trying to second guess what the other guy has determined (i.e., two steps removed). Assessors' determinations might align more closely, versus true essentiality, in the real world due to unconscious or conscious prejudice, coordination or conspiring between assessors.

By identifying and counting the essentiality determination agreements and disagreement between the two assessors, as well as counting the agreements and disagreement between the assessors and what is truly essential, I have been able to derive an additional scale, as also included in Exhibit 10. That translates Accuracy Rate A percentages into corresponding Consistency Rate C percentages. Both percentage figures are the same at 100% and 50%. Within that range they diverge with Accuracy Rates A higher than Consistency Rates C. The differences are independent of the Essentiality Rates TE (i.e., they do not vary with different TE percentages). I simply noted the changing values for the Consistency Rate C as I varied the Accuracy Rate A in my simulation model and added them in an additional column to my table.

The look-up table in Exhibit 10 can also be used to figure out what the True Essentiality Rate TE is by using comparisons of essentiality studies, such as from the EC Pilot Study's experiment. I model the relationship between A and C and also include corresponding values for them my look-up table. The EC Pilot Study report effectively states that a True Essentiality Rate TE=30% with an Accuracy Rate A=90% corresponds to a Found Essentiality Rate E1=E2 of 34%.[58] My modelling confirms this relationship. In in Exhibit 11, I have highlighted this 34% figure, along with the 36% figure from my example in Exhibit 9 and the 17%/18%[59] figure from my example in Exhibit 10.

A notable special case that is worth considering is with Consistency Rate C=50%. That is the same as my old favourite of determining essentiality with the toss of a coin. Under that condition, the Found Essentiality Rate is 50% because determinations are totally independent of whether the patents are truly essential.

[58] The EE Pilot Study report at page 80.

[59] The difference between 17% and 18% is in rounding and in the granularity of the table in Exhibit 11.

Exhibit 11: Look-up table of Found Essentiality Rates (E1 or E2) showing bias in these versus True Essentiality Rates (TE) at different Consistency Rates (C) and Accuracy Rates (A)

		True Essentiality Rate (TE)																				
C	**A**	**0%**	**5%**	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**	**45%**	**50%**	**55%**	**60%**	**65%**	**70%**	**75%**	**80%**	**85%**	**90%**	**95%**	**100%**
100.0%	0%	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%	45%	40%	35%	30%	25%	20%	15%	10%	5%	0%
90.5%	5%	95%	91%	86%	82%	77%	73%	68%	64%	59%	55%	50%	46%	41%	37%	32%	28%	23%	19%	14%	10%	5%
82.0%	10%	90%	86%	82%	78%	74%	70%	66%	62%	58%	54%	50%	46%	42%	38%	34%	30%	26%	22%	18%	14%	10%
74.5%	15%	85%	82%	78%	75%	71%	68%	64%	61%	57%	54%	50%	47%	43%	40%	36%	33%	29%	26%	22%	19%	15%
68.0%	20%	80%	77%	74%	71%	68%	65%	62%	59%	56%	53%	50%	47%	44%	41%	38%	35%	32%	29%	26%	23%	20%
62.5%	25%	75%	73%	70%	68%	65%	63%	60%	58%	55%	53%	50%	48%	45%	43%	40%	38%	35%	33%	30%	28%	25%
58.0%	30%	70%	68%	66%	64%	62%	60%	58%	56%	54%	52%	50%	48%	46%	44%	42%	40%	38%	36%	34%	32%	30%
54.5%	35%	65%	64%	62%	61%	59%	58%	56%	55%	53%	52%	50%	49%	47%	46%	44%	43%	41%	40%	38%	37%	35%
52.0%	40%	60%	59%	58%	57%	56%	55%	54%	53%	52%	51%	50%	49%	48%	47%	46%	45%	44%	43%	42%	41%	40%
50.5%	45%	55%	55%	54%	54%	53%	53%	52%	52%	51%	51%	50%	50%	49%	49%	48%	48%	47%	47%	46%	46%	45%
50.0%	**50%**	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%
50.5%	**55%**	45%	46%	46%	47%	47%	48%	48%	49%	49%	50%	50%	51%	51%	52%	52%	53%	53%	54%	54%	55%	55%
52.0%	**60%**	40%	41%	42%	43%	44%	45%	46%	47%	48%	49%	50%	51%	52%	53%	54%	55%	56%	57%	58%	59%	60%
54.5%	**65%**	35%	36%	38%	39%	41%	42%	44%	45%	47%	48%	50%	52%	53%	55%	56%	58%	59%	61%	62%	64%	65%
58.0%	**70%**	30%	32%	34%	36%	38%	40%	42%	44%	46%	48%	50%	52%	54%	56%	58%	60%	62%	64%	66%	68%	70%
62.5%	**75%**	25%	27%	30%	32%	35%	37%	40%	42%	45%	47%	50%	52%	55%	58%	60%	63%	65%	68%	70%	73%	75%
68.0%	**80%**	20%	23%	26%	29%	32%	35%	38%	41%	44%	47%	50%	53%	56%	59%	62%	65%	68%	71%	74%	77%	80%
74.5%	**85%**	15%	18%	22%	25%	29%	32%	36%	39%	43%	46%	50%	54%	57%	61%	64%	68%	71%	75%	78%	82%	85%
82.0%	**90%**	10%	14%	18%	22%	26%	30%	34%	38%	42%	46%	50%	54%	58%	62%	66%	70%	74%	78%	82%	86%	90%
90.5%	**95%**	5%	9%	14%	18%	23%	27%	32%	36%	41%	45%	50%	54%	59%	64%	68%	73%	77%	82%	86%	91%	95%
100.0%	**100%**	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%

Source: WiseHarbor

Areas of greatest interested are unshaded. Examples identified are highlighted.

Normalising results to eliminate bias in essentiality-rate determinations

One can envisage what might be proposed as "suitable statistical corrections" to fix to reduce or eliminate the statistical bias in determination that would force Found Essential Rates lower to align more closely with True Essentiality Rates.[60] This could be like the way exam results are normalised or staff are stack-ranked. Essentiality assessors could be asked to rate patents on a 1 to 10 scale on "how essential" they are, instead of the binary choice of essential or not essential. And then, results could be scaled so that, for example, only the 10% most essential patents are deemed essential.

While this could be the subject of another empirical study, I doubt such adjustments would help and might even make matters worse. The whole point of all this is that we do not know what the essentiality rate truly is for the universe of patents in the denominator, or for any subject company in the numerator. I suspect the agreement accuracy on which patents are essential would fall significantly with such an approach. Scaling-down the count of found-essential patents seems unlikely to improve accuracy on how many are found essential, let alone which ones.

Before doing that, it would be helpful to distinguish between the agreement rates (i.e., Consistency Rates and Agreement Rates) on essential versus not essential patents, so "corrections" could reflect those differences.

Tolerance to random variations due to sampling and inaccurate essentiality determinations

It is still rather unclear, even approximately where True Essentiality Rates TE and Accuracy Rates A lie. The EC's Pilot Study report at page 80 states:

> "***For illustration***, *assume that 30 percent of a sample of declared SEPs are actually essential and the likelihood of an incorrect assessment is 10 percent. Then the assessment yields, on average, a 34 percent essentiality rate, hence a relative overestimation by 13 percent" (emphasis added).*

The basis for these figures is unclear. As already indicated, my modelling and simulations results agree with that numerical relationship finding (see Exhibit 11), but those assumptions are questionable. In contrast, experts including Cooper and his co-authors find that True Essentiality Rates TE are only 12% for 4G LTE and 8% for 5G.[61]

If the aim is to determine top-down royalty rates with "tolerance" of ±15%, as a proportion of the base dollar value or percentage figure (e.g. $1.00 ± 15 cents) at the 95% confidence interval level, as I tentatively suggest, essentiality rates for the denominator of patent counts must also be determined to at least the same precision. This ignores any imprecision in determining the numerator, which would further increase the variability of overall determinations in top-down FRAND-rate setting.

With Accuracy Rates of A=90% (equivalent to Consistency Rates C=82%) and a True Essentiality Rate of TE=30%, a sample size of 200 would yield a tolerance of ± 21.8%, a sample of 500 would yield a tolerance of ±14.7% and a sample of 1,000 would yield a tolerance of ± 9.8%.

---

[60] The EC Study team seem to be suggesting that at page 81 "Such bias can be addressed in a future implementation by suitable statistical corrections".
[61] They also seem to implausibly imply that their determinations can be made faultlessly on small numbers of patents, if there is adequate time given for their assessments.

However, if Accuracy Rates of 90% can be achieved (equivalent to 82% Consistency Rates) with a True Essentiality Rate of only 20%, a sample size of 200 would yield a tolerance of ± 31.7%, and a sample of 1,000 would yield a tolerance of ± 13.8%.

So, a sample size of at around 500 is required, even if the True Essentiality Rate is 30%. However, I suspect the True Essentiality Rate may be rather lower than that, as Cooper et al. asset. Moreover, that rate seems likely only to be decreasing given the "skyrocketing" increase in SEP filings and declarations.[62]

More work to do

My simulation modelling is the first I am aware of that measures the combined effects of patent sampling and imperfect essentiality determinations on essentiality rate estimation accuracy. My analytical model and simulation is based on some figures made publicly available from the EC's Pilot Study experiment, as revealed in its report on that. This included the Consistency Rate figures of 75% (without claim charts) and 84% with claim charts.

It is unclear whether assessors are generally more consistent in finding non-essentiality than essentiality (or vice versa) than my simulations assume. However, in the very small number of instances where two different TCL experts in *TCL v. Ericsson* assessed the same patents for essentiality reveals different Consistency Rates on found essential patents than on found not essential patents.

Exhibit 12: Disagreements among friends in the TCL camp in TCL v. Ericsson

	Dr Ding (Concur IP)	Dr Kakaes	2G	3G	4G	TOTAL
Disagree	Essential	Not Essential	0	0	5	5
	Not Essential	Essential	2	3	6	11
Agree	Essential	Essential	5	10	15	30
	Not Essential	Not Essential	0	3	10	13
Total			7	16	36	59
Disagreement rate overall			29%	19%	31%	27%
Consistency Rate overall			71%	81%	69%	73%

As well as showing that different assessors (even while working in concert on behalf of TCL) disagree in a substantial 27% of their determinations overall (a Consistency Rate of C=100%-27%=73%), this small sample also suggests that the rate of agreement on which patents are essential (i.e., 30/59=51%) can be significantly different to the rate of agreement on which patents are not essential (i.e. 13/59=22%).[63] Adding these two percentages results in the overall Consistency Rate of 51+22=71%. And, with both assessors agreeing that most patents in this sample are essential, Dr

[62] An article by IPlytics published in July estimates over 150,000 declared patents for 5G alone since 2015. https://www.iam-media.com/ai-may-be-the-solution-skyrocketing-numbers-of-sep-declarations

[63] While this example shows that disagreements can be substantial, and it illustrates mechanics of conditional probabilities, the sample of patents is not representative of the universe of SEPs. These were all patents Ericsson had verified to be essential in litigation and for which it had produced claim charts. It is only to be expected that the essentiality rates found here were much higher for this small sample of patents than for the entire universe of patents, as determined by Dr Ding with assistance from Concur IP to be 37.3%.

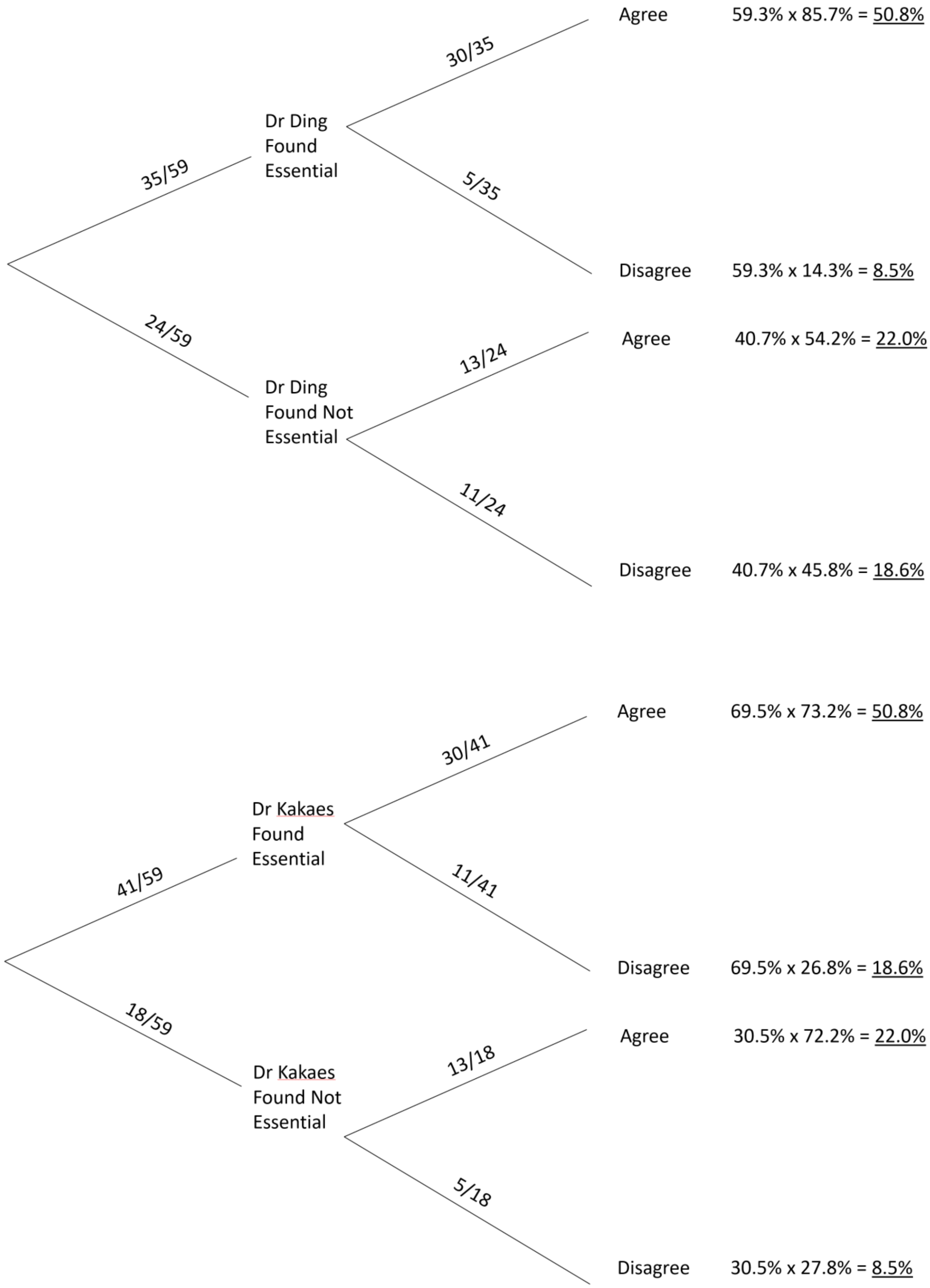
Ding’s and Dr Kakaes’ Found Essentiality Rates are (5+30)/59= 59% and (11+30)/59=69% respectively.

Exhibit 13 presents two conditional probability trees based on the figures above.

Modelling could be further developed to account for such differences with more detailed information on all determinations by two different assessors in larger samples. Modelling could also be used to deduce correspondingly different Accuracy Rates for found essential and found not essential patents versus truly essential and truly not essential patents.

This would be useful because, for example, with True Essentiality Rates asserted to be as low 10% by some experts, applying the same combined Accuracy Rate results in 9 times more false positives than false negatives. This creates the statistical bias in Found Essentiality Rates that my analysis reveals because only 1 out of 10 of false determinations nets off. Modelling a bifurcation in the Accuracy Rate, with some kind of accuracy rate figure for found essential patents and a different accuracy rate figure for found not essential patents— from empirical data— would enable us to see if biases differ, and if so, by how much. It could thus help us estimate True Essentiality Rates more precisely.

Exhibit 13: Conditional probability trees for essentiality determinations on patents assessed by both Dr Ding and Dr Kakaes in TCL v. Ericsson



Conclusions

Despite all the legal, economic and practical shortcomings in top-down analysis for FRAND-rate determination, it seems that use of the technique is not going away any time soon. Consequently, the counting of patents assessed to be standard-essential seems inevitable. Patent counts are also commonly used when determining royalty rates in comparable licenses. SEP counts are used there to derive the patent “portfolio strength ratio” required to “unpack” cross-licenses licenses and derive “one-way” licensing rates.

Essentiality checking is also proposed by the European Commission and others to improve transparency for prospective licensees.[64] But transparency is only legitimate if what is being revealed and counted is reasonably accurate. Otherwise it will likely do more harm than good.

I have highlighted concerns over many years about patent counting and its proposed institutionalisation.[65] This erroneously emphasises quantity over quality; with large numbers of patents that are declared as possibly standard essential, and are claimed to be standard essential, prevailing over far fewer than can be proven to be standard-essential. While in a free market parties in negotiation and in litigation should be at liberty to present whichever metrics and studies they wish, centralising essentiality checking and patent counting in some kind of ministry of patent counting, as proposed by the European Commission, is a dubious initiative.

Nevertheless, if patent counting is to persist, it is important to understand how imprecise SEP counts are and recognise the significant effects they have on accuracy and reliability in FRAND-rate determinations, particularly when using different and more exacting methods of determining patent essentiality in the numerator of top-down calculations.

It is abundantly clear there are far too many declared essential patents to assess all of them for essentiality. It would be too costly and take much too long.

Automated methods hold promise in helping with patent selection and in essentiality assessment, but AI is unproven in the latter. Before adoption or relying upon AI, it is vital that techniques are tested and verified with other methods.

Use of patent counting—with or without sampling—is an under-researched and impetuous leap of faith. Because it is impossible to find out for sure which patents are truly essential or even what proportion of them are essential (i.e. the True Essentiality Rate TE), helpful primary research is in comparing the essentiality-determination results of two different assessors working independently on the same set of representative patents. An EC Pilot Study experiment did that on a small sample of 205 patents.

However, results from such studies need careful interpretation because differences in determinations between independent assessors are not equivalent to the differences between an assessor’s determinations and what is truly essential.

Such comparisons and interpretations indicate that assessors tend to systematically inflate True Essentiality Rates. These over-estimates result from statistical bias with numbers of false positives (i.e., truly not essential patents being found to be essential) exceeding false negatives when true essentiality rates are rather less than 50%. Results also show that assessors’ determinations will tend

[64] https://ec.europa.eu/growth/content/transparency-sep-licensing-how-clarify-possible-exposure-upfront_en
[65] https://www.wiseharbor.com/wp-content/uploads/2017/05/Patent-Counting-article-for-IP-Finance-12-May-2017.pdf

to be in closer agreement with what is truly essential than their determinations will be in agreement with each other. Having a single metric for accuracy in essentiality determination implies and assumes false positives and false negatives are at the same rate. At low rates of true essentiality the number of false positives will thus greatly outnumber false negatives.

My analysis with simulations using empirical inputs also indicate that errors and disparities introduced by sampling and in imperfect essentiality assessments result in substantial essentiality rate inflation with statistical bias and variability in the essentiality rates determined. The problem is particularly acute if essentiality rates are relatively low, for example, at only around 10% or less.

I also find from my simulation modelling and analysis that:

1. The lower the True Essentiality Rate and the lower the rate of agreement among different assessors, the larger the differences will be between the essentiality rates determined by assessors and the True Essentiality Rate. For example:

    a. If True Essentiality Rates are 30% and two different assessors agree with each other on 75% of their determinations, they will tend to estimate 36% essentiality rates and be accurate in 85% their determinations.

    b. If True Essentiality Rates are only around 10% (e.g., for 4G LTE or 5G), as some experts plausibly argue, and if two assessors agree with each other on 84% of their determinations, they will tend to estimate 17% essentiality rates and be accurate in 91% of their determinations. That means there will be nearly as many false positives as correct determinations of essentiality.

2. Therefore, if True Essentiality Rates are at the lower end of expectations, for example, at around 10% or less, it is imperative assessors are highly accurate in their determinations, otherwise false positives will swamp their correct determinations and make their overall results meaningless.

The above constraints on sample sizes and overall accuracy rates might be relaxed somewhat if the rate of false positive determinations is found to be substantially lower than for false negatives. That is an empirical question requiring further work.

Encouraging assessors to grade more strictly would have the effect of increasing agreement on which patents are found not essential, and would reduce Found Essentiality Rates. However, how much that would reduce the accuracy in determining patents that are found to be essential is unclear but might be significant.

Sampling reduces the precision of SEP counts with increased variability, which is exacerbated by erroneous essentiality determinations. Sampling theory and simple simulations using results of patent-counting studies already undertaken—including several with sample sizes below a few hundred—reveal unacceptably large ranges in expected essentiality rate determinations (i.e. the percentage of declared-essential patents that are deemed to be essential)[66] at what various study authors regard as the "well accepted bound" of the 95% confidence level.[67] This variability is particularly large where patent essentiality rates are at low levels, such as at around 10%. For example, 10% ± 1.5% is actually ± 15% variability as a proportion of that 10% figure. Quantitative

[66] This is also called the essentiality ratio.

[67] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3752684 at page 4 and https://media.crai.com/wp-content/uploads/2020/09/16163609/Transparency_predictability_efficiency.pdf at page 38

analysis I undertook for this article measures the extent of diminutions, which should be properly and fully considered before sample sizes are set, and before the short cut of sampling is blindly adopted at all.

While there are no set bounds for the acceptably accurate range in determinations, I have considered a reasonable proportionate accuracy requirement for essentiality rate determination to be <± 15% (i.e., a 30% range for the determined essentiality rate as a proportion of the true essentiality rate) at the 95% confidence interval level. With expert opinion that True Essentiality Rates are more like 10% or even less than 30% or 40%, I conclude from my simulation analysis that samples including several thousands of patents per standard are required in top-down FRAND-royalty rate setting. For example, if the essentiality rate is only 10%, as some experts plausibly argue, a sample size approaching 3,000 declared-essential patents per standard, at the very least, would be required.

About this publication and its author

This article was originally published in the IP Finance blog on 30th September 2021.

Keith Mallinson is founder of WiseHarbor, providing expert commercial consultancy since 2007 to technology and service businesses in wired and wireless telecommunications, media and entertainment serving consumer and professional markets. He is an industry expert and consultant with 25 years of experience and extensive knowledge of the ICT industries and markets, including the IP-rich 2G/3G/4G mobile communications sector. His clients include several major companies in ICT. He is often engaged as a testifying expert witness in patent licensing agreement disputes and in other litigation including asset valuations, damages assessments and in antitrust cases. He is also a regular columnist with RCR Wireless and IP Finance – "where money issues meet intellectual property rights."

The author can be contacted at WiseHarbor. His email address is kmallinson@wiseharbor.com and you can also follow him on Twitter at http://twitter.com/WiseHarbor.